ACTIVITY RECOGNITION FOR AMBIENT SENSING DATA AND RULE BASED ANOMALY DETECTION

A THESIS SUBMITTED TO
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
KOCAELİ UNIVERSITY
BY

EHSAN SEYEDKAZEMI ARDEBILI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

KOCAELİ 2022

ACTIVITY RECOGNITION FOR AMBIENT SENSING DATA AND RULE BASED ANOMALY DETECTION

A THESIS SUBMITTED TO
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
KOCAELİ UNIVERSITY

BY

EHSAN SEYEDKAZEMI ARDEBILI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

Prof.Dr. Kerem KÜÇÜK
Supervisor, Kocaeli University .....

Prof.Dr. Cüneyt BAYILMIŞ
Jury member, Sakarya University .....

Assoc.Prof.Dr. Süleyman EKEN
Jury member, Kocaeli University .....

Thesis Defense Date: 13.01.2022

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- Bu tezin bana ait, özgün bir çalışma olduğunu,
- Çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun olduğunu,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı,

beyan ederim.

■ Bu tez çalışmasının herhangi bir aşaması hiçbir kurum/kuruluş tarafından maddi/alt yapı desteği ile desteklenmemiştir.

□ Bu tez çalışması kapsamında üretilen veri ve bilgiler ..... tarafından ..... no'lu proje kapsamında maddi/alt yapı desteği alınarak gerçekleştirilmiştir.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

.....

Ehsan Seyedkazemi Ardebili

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanımı bana ait olacaktır.

Tezin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim kurulu tarafından yayınlanan ***"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*** kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

□ Enstitü yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir.

□ Enstitü yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir.

■ Tezim ile ilgili gizlilik kararı verilmemiştir.

.....

Ehsan Seyedkazemi Ardebili

PREFACE AND ACKNOWLEDGMENT

Firstly, I would like to express my gratitude and reverence to ALLAH ALMIGHTY, who gives me strength, ability, knowledge, and patience to continue my studies as well as the perseverance to complete my master's degree satisfactorily. I believe that without his blessings, it would have been impossible.

I wish to express my sincere gratitude and warmest thanks to my supervisor, Prof. Dr. Kerem Küçük, for his continuous support, guidance, motivation, and supervision throughout my studies. His professional approach and practical knowledge helped me to complete my master's degree in the best possible way. It wouldn't be possible without his significant support.

I take pride in acknowledging the insightful guidance of Assoc.Prof. Dr. Suleyman Eken. I'm truly privileged to study under his guidance. I salute his full support, invaluable advice, motivation, and suggestions during my studies.

I am also grateful to Prof. Dr. Adnan Kavak for his invaluable time, comments, suggestions and contributions to my research.

Besides my professors, I am really thankful to all my friends. I am also very thankful to all the academic personnel of the Computer Engineering Department at Kocaeli University, Turkey.

Last but not least, I am obliged to my Parents for their unconditional love, support, and prayers. I'm very much thankful to my brothers and sisters for their love and support.

January – 2022 Ehsan SEYEDKAZEMI ARDEBILI

CONTENTS

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI ii
PREFACE AND ACKNOWLEDGMENT iii
CONTENTS iv
LIST OF FIGURES v
LIST OF TABLES vi
LIST OF ABBREVIATIONS vii
ÖZET viii
ABSTRACT ix
1. INTRODUCTION 1
2. BACKGROUND AND MOTIVATION 6
    2.1. Related Works 6
    2.2. Reviewing Objectives of the Studies 16
    2.3. Reviewing the Provided Methods 19
    2.4. Datasets Classification 20
3. HUMAN ACTIVITY RECOGNITION WITH AMBIENT SENSING DATA 22
    3.1. Designing the Approach and Traditional Methods 22
    3.2. Principle One – Privacy of Residents 31
    3.3. Principle Two - The Possibility of Severe and Daily Irregularities 32
    3.4. Principle Three - Using Light and Simple Calculations 32
    3.5. Description of the Anomaly Detection Approach 34
4. EXPERIMENTS AND RESULTS FOR HUMAN ACTIVITY RECOGNITION 36
    4.1. Designing the Approach 36
    4.2. Description of the Rule-Based Concept 37
    4.3. Description of Sub-NNC 40
    4.4. Definition of Daily-NNC and Main-NNC 42
    4.5. Sub-NNC Based on Statistical Techniques 46
    4.6. Sub-NNC Based on Machine Learning Techniques 48
    4.7. Experimental Setup for Investigating the Performance 50
    4.8. Performance Metrics 51
    4.9. Performance Results 53
    4.10. Aras Dataset 54
5. EFFORTS TO INCREASE PRODUCTIVITY AND ACCURACY 56
    5.1. Examining Possible Methods to Improve 56
    5.2. New Models' Definition 60
    5.3. Deep Learning 63
6. CONCLUSIONS 66
    6.1. Assumptions and Limitations 66
    6.2. Future Studies 67
    6.3. Summary 68
REFERENCES 74
PERSONAL PUBLICATIONS AND ARTICLES 78
BIOGRAPHY 79

LIST OF FIGURES

Figure 4.1. Photos from inside a smart home……...…...39

LIST OF TABLES

Table 2.1. Summary of studies related to this field.....6
Table 4.1. Description of sub-NNCs that are based on statistics.....45
Table 4.2. Description of sub-NNCs that are based on machine Learning.....46
Table 4.3. Classification of predictions.....52
Table 4.4. Test results of the top four machine learning algorithms.....54
Table 5.1. Test results of the top four machine learning algorithms.....61

LIST OF ABBREVIATIONS

ARAS	: Activity Recognition with Ambient Sensing
BMS	: Building Management System
DL	: Deep Learning
IoT	: Internet of Things
ML	: Machine Learning
NN	: Neural Network
NNC	: Number of Normal Changes
SHMC	: Smart Home Management Center

ORTAM ALGILAMA SENSÖRLERİNDEN ELDE EDİLEN VERİLERDEN AKTİVİTE SINIFLANDIRMASI VE KURAL TABANLI ANOMALİ BELİRLEME

ÖZET

Bu yüksek lisans tezinde, ortamda bulunan sensör verilerinden kural tabanlı anomali tespiti üzerinde çalışılmıştır. Günümüzde akıllı eve olan ilginin artması ve yeni teknolojilerin ortaya çıkması, insanların yaşam tarzlarında meydana gelen değişiklikler, artan nüfus artışı, müstakil evlerin sayısında da artış gibi farklı nedenler ile akıllı evler yoğun olarak kullanılmaya başlanmıştır. Bu konudaki üç önemli başlık göz önüne alınmaktadır.

Bu konuda yapılan çalışmaları detaylı bir araştırması sunulan bu tezde yapılmış çalışmalar detaylı bir şekilde değerlendirilmiştir ve bu konunun öneminden dolayı birçok makale ve araştırma çalışması ile karşılaştırılmıştır. Bu çalışmada, genel olarak çalışmaları önce veri setlerinin türüne göre, daha sonra problemi çözmek için önerdikleri yaklaşıma göre sınıflandırılmıştır. Daha sonra genel olarak mevcut yöntemlerin avantaj ve dezavantajları gözden geçirilmiştir ve tüm bu adımlardan sonra anormal durumu en doğru ve en kısa hesaplama süresi ile belirlemek için bir yöntem önerilmektedir.

Bu tezde, Normal Değişikliklerin Sayısı (Number of Normal Changes, NNC) konseptini temel alan bir yöntem önerilmektedir. Bu yaklaşım, her ev sakininin davranışlarını gözden geçirerek, her bir kişinin farklı zaman ve durumlardaki davranış kalıplarını bulmaya çalışır. Ev sakinlerinin davranışlarındaki anormallikleri bulmak için, her bir kişideki davranış değişikliklerinin kapsamını tespit edildiği bir dizi kurallar bütünü sunulmaktadır. Önerilen yöntemde herhangi bir ek akıllı cihaz takmaya veya taşımaya gerek bulunmamaktadır ve evde özel ek ekipman kullanılmasına gerek yoktur. Bu yöntem minimum sensör ve işlemci ile uygulanabilir durumdadır. Önemli olan nokta, tüm bu aşamalarda kişilerin mahremiyetini asla ihlal edilmemesidir. Bu yüksek lisans tezinin sonunda, tüm bu çalışmaların, sakinleri, özellikle evini seven ve huzurevine gitmek istemeyen yaşlılar için akıllı evleri güvenli ve keyifli bir ortama dönüştürebileceğini gösteren sonuçlar sunulmaktadır.

Anahtar Kelimeler: Akıllı Ev, Anormallik Algılama, İnsan Aktivitesi, Makine Öğrenimi, Nesnelerin İnterneti.

ACTIVITY RECOGNITION FOR AMBIENT SENSING DATA AND RULE BASED ANOMALY DETECTION

ABSTRACT

In the first part of this master's thesis, studying the reasons for increasing attention to smart homes shows the fact that there are different reasons behind this; for instance, the emergence of new technologies and changes in people's lifestyles, as well as the increase in the number of single-family homes, but here three main reasons will be presented.

In the next step, the previous studies will be evaluated in detail first, and due to the importance of this topic, many articles and research will be examined. Furthermore, a variety of methods and approaches will be discussed. In this research, studies are generally classified based on the type of datasets first and then based on the approach they have proposed to solve the problem. After that, the advantages and disadvantages of the existing methods are reviewed in general which will be followed by finding a way to recognize the abnormal state with the high accuracy and low latency (the fastest computation time).

A method based on a concept called NNC will be proposed. It takes an overview of each resident's behavior, and tries to find each individual's behavioral patterns at different times and situations. In order to find abnormalities in the behavior of the house occupants, a series of rules will be introduced in an attempt to find the extent of behavioral changes in each person. There is no need to wear or carry any additional smart devices in the proposed method, and there is no need to use special additional equipment at home. Moreover, this method can be implemented with a minimum of sensors and processors. The critical point is that people's privacy will never be violated in all these stages. The most tremendous respect for people’s privacy is taken into consideration. And at the end of this master's thesis, it is hoped that all the studies will be able to turn smart homes into a safe and pleasant environment for their residents, especially for the elderly who love their home and do not want to go to a nursing home.

Keywords: Smart Home, Anomaly Detection, Human Activity, Machine Learning, Internet of Things.

1. INTRODUCTION

Home is the most important place in every person's life and one of the places where people spend most of their lives. The first man-made building is a house, and from the first days of the emergence of human societies, human beings have been working tirelessly to build better and more comfortable homes for themselves.

At first, the house may have been just a shelter to protect people from adverse weather conditions, but today, it has taken on a new definition (after so many years). In human societies, we see all kinds of houses that are improving every day and new facilities are added to them.

Houses was becoming more advanced and smarter as the time went by. Perhaps the world's first sparks of BMS were formed in the late 1960s. It is clear that the smart home at that time was different from the concept we use today. Moreover, perhaps the first steps in making homes smarter were the advent of equipment that contributed to human comfort. We can even consider the first washing machine or the first home iron as one of the first steps in making houses smart.

In general, we can say today's ‘smart home’ means a home that has a wide connection between all the used equipment, facilities, and the possibility of controlling them from any place and at any time. ‘Smart home,’ ‘smart building,’ or ‘automated building’, which have very close definitions, are trying to describe a new and expanding generation of homes. Homes at which their most basic level can have features such as doors, electrical curtains, and a security system. In the middle and the next generations, smart homes have basic management systems such as intelligent lighting and temperature control systems or smart audio-video system; in addition, at the advanced level, it has advanced systems of energy consumption management system or has a central smart management system of the house, which performs routine tasks completely automatically, issues control commands for various tasks, and has the ability to communicate with people living in the house.

Therefore, according to the initial definition, smart homes use a central controller enabling the communication between all existing devices and equipment; thus, with such

a management center, which is like the brain of the house, a new lifestyle is being experienced in these new smart homes.

Smart homes have been started around the world since 1990, but it was not important until 2000 that the number of smart homes increased dramatically, and from 2010 onwards, the smart home was trendy among different societies. There must have been a variety of reasons for this sharp increase, but in this master's dissertation, these reasons are categorized into six main groups:

a- The emergence of the Internet of Things on a global scale and the beginning of the expansion of the Internet of Things in all environments strengthened and created the next infrastructure for the smart home.

b- At the beginning of the advent of smart buildings in the world, having a smart building was very expensive. The cheapness and availability of hardware and software needed for smartening up changed the definition of a smart home from an almost unnecessary luxurious product.

c- Increasing people's attention to management and saving time and energy itself in recent years, given that energy reserves in human hands are in danger of running out, as well as the importance and value of the concept of time for all. Homes that manage their own energy consumption have found a special place. Therefore, having houses with lower energy consumption can reduce the staggering costs in the future. Needless to say, saving in a smart home does not mean consuming less, and you do not need to reduce the consumption of residents, but it is consumed properly, and by installing smart equipment in homes, no additional energy will be consumed, and energies will be used purposefully.

d- Significant advances in the field of artificial intelligence and the emergence of fully functional and useful devices in smart home environments have made people inclined to this type of home and have made their homes smarter to take advantage of these devices.

e- One of the reasons that may not be very obvious at first glance, but in a closer look one can find that home builders are increasing attention to smart homes, and that is because smart homes are attractive to home buyers, and thus by making the house smarter, it will be sold both faster and with a higher price.

f- Governments have supported smart homes for a variety of reasons. As an example, one can refer to the fact that governments are aware of the extent to which these homes can reduce energy consumption in their countries; therefore, according to international standards, they have attempted to prepare rules and regulations for the construction of smart homes in their own countries, and to support this type of house in the construction and renovation of houses. This issue is so important that in some societies it is predicted that such a process can reduce the energy consumption of a country up to 30%.

After drawing on the definition of ‘smart home’ and the reasons behind the importance and attention to the issue, in this master's dissertation, the subject will be focused on in more detail, and components of the smart home will be divided into three parts: hardware, software, and Smart Management Center.

In the first part, the hardware, which consists of a new generation of hardware such as touch keys or a new generation of sensors and advanced cameras and smart home appliances, will be discussed. In the software section, interfaces and control programs which are used to control the hardware will be discussed, and the interface is between the hardware and the occupants of the house, and information is exchanged in its context. Perhaps the simplest example of this section is the application that is installed on the phone of residents of the house to issue control commands such as turning on a light bulb.

The last and most important part of a smart home is the smart management center, the place where control commands are issued. Although this part can be considered as a subset of the software section, due to the importance of this section, it is examined separately. All automatic commands are managed in this section and like a human brain which manages all the organs of the body, this section manages all the components of the house, and all the devices and all the parts of the house are under the supervision of this section.

To make a house smart can have different meanings and definitions, but perhaps one can say anything which happens spontaneously and without human intervention can be considered as an example of the word ‘smart.’ However, in a more comprehensive definition of smart homes, it has been pointed out that all devices and consumers in smart buildings perform all their activities or at least part of their activities automatically and without human intervention.

For example, when you enter a part of the house, the lamp of the same part turns on and it turns off when you leave, or keeps the home's heating and cooling system at the right temperature set by the homeowner. Hence, intelligence is able to take place at different levels. As a result, making a home smart can refer to all the services and products which perform or display an action or message in the home environment without the physical intervention of the homeowner. indeed, we know that not every action is considered spontaneous, and these things happen mechanically, i.e. according to the needs and conditions of the time and place of the house.

To have a correct spontaneous event or a correct management decision, the first requirement of a smart home management center is a correct understanding of home conditions. As the human brain examines the situation before making any decisions, in a smart home, before making any decisions, the home management center should check the condition of the house, i.e. what is the condition of the house, and whether it is normal or abnormal. As it is known, the most important element of any home is human beings present in it, and the most influential things happen to the occupants of the house. Therefore, in this master's thesis, the residents' behavior of the house will be examined and abnormalities will be diagnosed.

After examining the importance of the issue, the existing methods and previous studies along with an accurate and general classification of the presented methods will be dealt with. Then, it is examined how the existing datasets obtained their information in order to implement those methods. After this step, advantages and disadvantages of the existing methods will be generally reviewed, and in the main part of this study, an approach will be presented. In this approach, it has been considered that it does not impose any additional costs on the 'smart home' and does not require a specialized device.

In some methods, due to the type of datasets which have been selected, the privacy of the residents is not respected, or in some cases, the presence of the camera in the bedrooms does not make the residents feel comfortable. In this thesis, due to the importance of human privacy, the issue will be dealt with carefully and respectfully which is hoped that the privacy of individuals will not be violated at any stages throughout the research.

At the end of this master's thesis, it is hoped again that all these studies make smart homes a safer and more enjoyable environment for their residents, especially for the elderly who

love their dwelling deeply. In addition, it has been endeavored to help improve research in this area and highlight the importance of having smart assistants in smart homes to detect human behavior. In an attempt to better understand what is being done in this thesis, a plan of the process and the taken steps have been outlined. Therefore, the entire thesis is summarized in the following seven steps:

a - Introduction and definition of the topic

b - Detailed review of related works and motivation

c - Explaining the basic principles of the thesis

d - Introducing and describing the proposed method

e - Evaluating the performance and accuracy of the proposed method

f - Efforts improving the proposed method

g - Conclusion and summary

2. BACKGROUND AND MOTIVATION

2.1. Related Works

To examine previous related works in the field of human behavior detection and abnormality detection, first, the reason for the importance of this issue is reexamined from another perspective which is different from the viewpoint in the introduction section on smart homes. Furthermore, the goals are discussed and classified in previous studies, and then the approaches proposed in previous researches are reviewed in general. Moreover, how the datasets used in those studies are evaluated, and then advantages and disadvantages of existing approaches are subjected to scrutiny. It should be mentioned that all the efforts have been to review all the articles in the field of recognizing human behavior and diagnosing abnormalities throughout the thesis.

Table 2.1. Summary of studies related to this field

No.	Author year	Dataset sample	Paper Title	Objective	Contributions	Methodology
1	(Alemdar 2014)	ARAS Dataset	Multi-Resident Human Behaviour Identification in Ambient Assisted Living Environment	human behaviour recognition in ambi- ent assisted living environments where multimodal interaction community is essential.	(i) Dataset collection from two real smart homes (ii) Multi- resident behaviour moni- toring and performance evaluation metrics	factorial hidden Markov models (FHHMM) and scalable machine- learning methods
2	(Parvin et al. 2018)	Task Model Simulator - In ambient Assisted Living (AAL) scenarios	Real-Time Anomaly Detection in Elderly Behavior with the Support of Task Models	They propose an online algorithm that detects anomalies at real-time	Its behaviour described according to two main phases: (i) analysis and modeling the user behaviour. (ii) analysis and modeling the user behaviour.	They used Context Manager (CM) and implemented a Task Model Simulator(TM S)

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
3	(Yachir et al. 2019)	They collected 500 records for each class	A Deep Learning and Multimodal Ambient Sensing Framework for Human Activity Recognition	Human Activity Recognition (HAR)	they distinguished six classes of elementary actions or postures namely: sitting, standing, running, walking, walking upstairs, and walking downstairs.	multiclass Support Vector Machine and Convolutional Neural Networks
4	(Deep et al. 2019)	On a variety of types Self-collected Publicly available	A Survey on Anomalous Behavior Detection for Elderly Care Using Dense-Sensing Network	review of existing articles and methods on Anomalous Behavior Detection that are suitable for the elderly	overview of the key challenges in current activity and anomaly detection technology	Dense Sensing Network Based Anomaly Detection - Vision Based Anomaly Detection Techniques - Wearable Sensors Based Anomaly Detection Techniques
5	(Diraco et al. 2019)	specific synthetic datasets are generated, which include activities of daily living, home locations and vital signs. HMM is used for data generation	Deep Learning and Machine Learning Techniques for Change Detection in Behavior Monitoring	compare state-of-the-art machine learning and deep learning approaches suitable for detecting early changes in human behavior	Abnormal Behavior Detection (Category): 1 Supervised detection – 2 Semi-supervised detection – 3 Unsupervised Detection	Machine learning: 1:SVM – 2: One-class support vector machine (OCSVM) – 3:K-means clustering (KM) Deep learning: 1:Convolutional neural network (CNN) – 2: Stacked auto-encoders (SAE)– 3:Convolutional auto-encoder (CAE)

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
6	(Makonin et al. 2012)	They conceptually model smart home as a complex ecosystem involving the occupants, physical and operational components of home, external and internal context, and dependencies between these.	a Smarter Smart Home: Case Studies of ambient Intelligence	the goal is to help occupants optimize home use through computational interventions without imposing undue technological complexity, effort, or inconvenience.	Smart homes are often difficult to manage (Even for technologically passionate residents) In general, efforts are being made to solve this problem	These interventions comprise a combination of information and interaction design; automation and control; adaptive intelligent agents; and distributed, ambient sensing.
7	(Jethananda ni et al. 2019)	ARAS dataset	Multi-Resident Activity Recognition using Multi-Label Classification in Ambient Sensing Smart Homes	The task of multi-resident activity recognition.	They evaluate the developed model of Classifier Chain with K-Nearest Neighbor as base classifier	Classifier Chain Method
8	(Natani et al. 2019)	ARAS dataset They used GANs to generate more data	Deep Learning for Multi-Resident Activity Recognition in Ambient Sensing Smart Homes	prediction of human activities using deep learning algorithms	They have used different variations of RNN on the dataset and measured their performance with fewer data and more data and with data generated with GAN.	(i) LSTM (Long short-term memory) networks are RNNs, adequate for learning long-term dependencies. (ii) GRU (Gated Recurrent Unit) is a special type of in RNN used to tackle the vanishing gradient problem of basic RNN.

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
9	(Yang et al. 2014)	BLEMS project Collection of sensor data and ground truth occupancy data in rooms 3 & 4 was carried out concurrently for 20 consecutive days	A systematic approach to occupancy modeling in ambient sensor-rich buildings	Aims to optimize the building's energy consumption while maintaining occupant comfort through computing, sensor monitoring, and online control	This paper evaluates occupancy modeling (both binary detection and multi-class estimation) using twelve ambient sensor variables. Performance of six ML techniques is evaluated in both single-occupancy and multi-occupancy offices	The study carried out a systematic evaluation of five different algorithms, namely SVM, ANN, NB, TAN, KNN, and DT
10	(Ertan 2013)	Daily living activities of residents were collected for 30 full days	Human activity recognition using wireless ambient sensor networks	With continuous and long term monitoring, daily routines of the residents can be inferred and any abnormal situation which can be an indicator of a disease can be detected.	They elaborate the system architecture of the wireless sensor network, Share the experiences obtained during the data collection process, and the results of the classification	. HMM, Decision Tree, k-NN and different feature sets for performance comparisons
11	(Fernandez -Carmona et al. 2017)	RGB-D and domotic data containing data entries from 37 different domotic sensors	Entropy-based abnormal activity detection fusing RGB-D and domotic sensors	New solution to combine local activity levels detected by a single RGB-D camera with the global activity perceived by a network of domotic sensors	The activity monitoring system in ENRICHME exchanges information between the domotic sensor network and the robot through the Ambient Intelligence Server (AIS).	Markov Logic Network (MLN) Definition a Markov network using a finite set of constants applied over clauses in a MLN.
12	(Igwe et al. 2018)	ARAS dataset	Activity Learning and Recognition Using Margin Setting Algorithm in Smart Homes	to introduce an accurate pattern recognition and activity learning method	(i) home automation (ii) capable of identifying trends and patterns a smart environment (iii) make effective decisions	The k-means clustering algorithm The CascadeSimple KMeans Margin Setting Algorithm

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
13	(Alemdar 2015)	ARAS dataset	Human activity recognition with wireless sensor networks using machine learning	Recognizing human behavior in an automated manner is essential in many ambient intelligence applications such as smart homes, health monitoring applications and emergency services.	(i) collecting publicly available benchmark datasets, (ii)proposing a methodology for incorporating a hierarchy into the model that is tailored for various activities individually, (iii)improving the ways of evaluating different approaches and models considering the domain specific needs,	FHMM for modeling two separate chains Nonlinear Bayesian tracking for decomposing the observation space
14	(Belapurkar et al. 2018)	I) Image, audio, video II) Data from various wearable sensor III) Data from ambient sensors	The Case for Ambient Sensing for Human Activity Detection	They make a case where ambient sensing (sensors that collect data representing only environmental changes, such as temperature, lighting, etc.) can be used to detect human activities	(i) Multimedia-based activity detection (ii) Wearable-based activity detection (iii) Ambient sensing-based activity detection	Support vector machines, Bayesian networks Random forests, Decision trees, Neural networks (SVM, ANN, DT, BN)
15	(Yoshimizu et al. 2014)	Collected data (Chair Sensor, Pen Sensor, Cup Sensor, Camera..)	Multi-sensor-based Ambient Sensing System for the Estimation of Comfort/Discomfort to Lighting Condition During Desk Work	A multi-sensor-based ambient sensing system is proposed for estimating the user's response to the lighting condition during desk work.	the recognition rate of the user's comfort/ discomfort on a two-point scale is 91% when selecting a suitable combination of ambient sensors.	Learning Phase: A->B->C-> Estimate parameter for SVM -> model training Recognition Phase: A->B->C-> Recognize Using SVM A) Extract Feature Values form each sensor data B) Normalize the feature values C) Generate feature vectors

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
16	(Khan and Roy 2017)	Three public datasets: i)HAR ii)Daily-AndSports iii)MHealth	TransAct: Transfer Learning Enabled Activity Recognition	Activity recognition using smartphone	They compare the performance of TransAct model with existing methodologies such as including Random Forest (RF), Decision Tree (DT) and Transfer Boost (TB).	the Instance based Transfer Boost algorithm with k-means clustering
17	(Oguntala et al. 2019)	Mock Room	SmartWall: Novel RFID-Enabled Ambient Human Activity Recognition Using Machine Learning for Unobtrusive Health Monitoring	They present a novel RFID-enabled data sensing approach for unobtrusive human activity recognition via the SmartWall.	The proposed algorithm based on the multivariate Gaussian via maximum likelihood estimation is used to learn the features of the human activity model. (The twelve sequential and concurrent experimental evaluations are conducted in a mock apartment environment)	Multivariate Gaussian distribution (They define the mean and covariance function to describe the distribution of the sampled data)
18	(Alemdar et al. 2013)	ARAS dataset	ARAS Human Activity Datasets In Multiple Homes with Multiple Residents	They present the details of ARAS (Activity Recognition with Ambient Sensing) human activity recognition datasets	The research efforts on smart homes mainly focus on 2fields. I) studies where hundreds of sensors and sensor equipped home appliances are deployed in a smart laboratory houses II) studies focus on human activity detection for health status monitoring	Hidden Markov Model (HMM) (for modeling the activities and sensor data since it is well suited for sequential data)

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
19	(Parvin et al. 2019)	The events received via the Context Manager. They created the task model which represents the elderly routine daily activities with the previously existing ConcurTask-Trees (CTTE) tool.	Personalized Real-time Anomaly Detection and Health Feedback for Older Adults	The central objective of our method is assisting older adults to maintain functional ability to do their daily routines and live as independently and healthy as possible.	The proposed solution models the users' daily routine using a task model specification and performs the online activity recognition through a Context Manager, which detects relevant contextual events occurring in their lives.	Mamdani-type Fuzzy Rule-Based System They used a fuzzy rule-based system to decide about the degree of the abnormality detected which results in identifying the level of intervention needed for the specific user. (Online and Personalized Anomaly Detection: Activity recognition aims to identify elderly activities based on a series of sensor readings. so they present the rule-based approach which extracts the activity from the simple events.)
20	(Ozen 2013)	Their testbed environment	Careact: an adaptive and context aware framework for anomaly detection in ambient assisted living	An ambient assisted living system which monitors daily lives of inhabitants for anomaly detection in real-time	They main contribution arises from two facts where the former is that no existing system in the literature supports such features in an all-in-one manner, and the latter is that CAREACT applies telemonitoring concepts to multimodality of both sensory and ADL recognition systems.	CAREACT: CAREACT is a framework under which various sensory and ADL recognition systems coexist and the data of those systems are fused by a pattern detection engine to deduce anomalies and to automatically create appropriate reactions on such anomalies in real-time. Rule Based Pattern Detection

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
21	(Stavropoulos et al. 2020)	GAADRD Clinical application lab	Semantic event fusion of computer vision and ambient sensor data for activity recognition to support dementia care	This paper presents such a framework that integrates both a computer vision component and heterogeneous sensors with unanimous semantic representation and interpretation.	They presents a framework build upon the intersection of complimentary fields of research, namely ambient intelligence, semantic web technologies, computer vision and ambient assisted living.	SI + HAR Semantic Interpretation (SI) The visual activity recognition component in the proposed framework, namely the human activity recognition (HAR) component, is based on image data utilizing computer vision techniques.
22	(Nef et al. 2015)	Dataset was collected in 10 homes of 10 healthy volunteers of various ages during 20 days each.	Evaluation of Three State-of-the-Art Classifiers for Recognition of Activities of Daily Living from Smart Home Ambient Data	They describe a technology that allows the ground truth collected in one smart home to train activity recognition systems for other smart homes	They compared the classifier performance by dividing raw sensor data into fixed-length time slots and classified each slot based on whether the sensors fired within that slot.	I)Naive Bayesian (NB) II)Support vector machine (SVM) III)Random forest (RF)
23	(Ang et al. 2016)	They deployed one Z-Wave Aeon Multi Sensor, one SmartThings SmartSense Open/Closed Sensor, one SmartThings SmartPower Outlet and one Netatmo Urban Weather Station.(for two weeks)	Human Occupancy Recognition with Multivariate Ambient Sensors	This paper aims to identify which ambient sensor is the most dominant in recognizing human presence.	Three experiments : (normalised raw data, normalised raw data was aggregated at ten minute intervals, the data is segmented into four temporal segments (morning, afternoon, evening, night))	i) Multi-layer Perceptron ii) Gaussian Processes(RBF) iii) SVM iv) Random Forest v)Naïve Bayes

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
24	(Bhargava et al. 2017)	Data collected outside their laboratory using the wearable device	Fog-centric localization for ambient assisted living	They present the design of our low-cost WSN system for mobility monitoring and outdoor localization of Alzheimer's patients.	Their system consists of an activity tracker and gateway node, and relies on self-tracking and sensor based analytics to perform autonomous, real-time localization as the user moves around an application environment.	on-board analytics approach along with the IEM algorithm that is used for on-board activity recognition
25	(Ghayvat et al. 2019)	collected from four elderly houses. (Most of the research work does not use heterogeneous unobtrusive sensing in an uncontrolled environment. Moreover, they use one or two home datasets.)	Smart Aging System: Uncovering the Hidden Wellness Parameter for Well-Being Monitoring and Anomaly Detection	This research work aims to apply lifestyle monitoring in an ambient assisted living (AAL) system by diagnosing conduct and distinguishing variation from the norm with the slightest conceivable fake alert.	The Wellness Model is an innovative methodology that classifies events inside definite temporal thresholds of activity initiation.	Wellness system (The Wellness Model processes data sampled at diverse rates and multi-modality sensory data with minimal and ordinal standards structured in hierarchies, not like numerous machine learning procedures that deal only with numeric entities.)
26	(Shelke and Aksanli 2019)	The dataset collected in their lab (non-contact, low-resolution infrared thermal sensors)	Static and Dynamic Activity Detection with Ambient Sensors in Smart Spaces	They demonstrated the potential use of low resolution thermal sensors for static and dynamic human activity detection in smart environments.	Low output resolution of these sensors enables real-time detection due to less computational overhead and mitigates privacy concerns since user identification is almost impossible.	1) Logistic regression 2)Naive Bayes 3) SVM 4) Decision tree 5) Random Forest 6) Neural network

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
27	(Aran et al. 2016)	real-life sensor data collected from 40 different households of elderly people (collected in the context of a Swiss project. The collected data includes 45 unique subjects, covering a total of 104 days.)	Anomaly detection in elderly daily behavior in ambient sensing environments	Their objective is to automatically observe and model the daily behavior of the elderly and detect anomalies that could occur in the sensor data.	Detect anomalies and significant changes in the behavior, with the assumption that these anomalies may signal health related problems. In the next sections, they present the sensing environment and the details of the dataset.	statistical model for summarizing behavioral (They define anomalies as events, which are different than the subjects' past behavior, as modeled by the behavior model defined)
28	(Wang et al. 2018)	The experiment test bed is a graduate student office located in an institutional building. The office has a gross area of about 200m2 and had 25 long-term residents during the experiment period.	Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings	compared three popular machine learning algorithms, including KNN, SVM, and ANN, combined with three data sources, to optimize the occupancy models' performance in various scenarios.	The results indicate that, comparing with independent data sources, the fused data set does not necessarily improve model accuracy but shows a better robustness for occupancy prediction.	i) K-Nearest Neighbors ii) Support Vector Machine iii) Artificial Neural Network
29	(Ghahramani et al. 2018)	Part of the US General Services Administration's environmental sensing project, office workers in a variety of roles were recruited in four Federal buildings for an observational study. (221 participants had valid work-time data collected)	Learning occupants' workplace interactions from wearable and stationary ambient sensing systems	they proposed a framework for learning occupants' interactions based on stationary and wearable sensing arrays	The framework uses recursive feature elimination method to select ambient sensing technologies which provide the optimal information gains and prevent overfitting.	Random Forests

Table 2.1. (Cont.) Summary of studies related to this field

No.	Author (year)	Dataset sample	Paper Title	Objective	Contributions	Methodology
30	(Otebolaku et al. 2020)	Data consists of those obtained from a combination of initial and ambient sensors	Deep Sensing: Inertia and Ambient Sensing for Activity Context Recognition using Deep Convolutional Neural Networks	They proposed to investigate the class imbalance problem of activity context recognition using deep convolutional neural networks to develop an effective and data-driven activity context recognition model using a combination of signals from sensors.	They propose to augment the time series signals from inertia sensors with signals from ambient sensing to train DCNN models (Recognition Accuracy using both inertial and ambient sensing data)	Deep Convolutional Neural Networks (Activity context recognition based on multivariate time series signals obtained from mobile devices in unconstrained conditions is naturally prone to imbalance class problems.)
31	(VandeWeerd et al. 2020)	USF's CREATE Health lab and The Villages (Florida were selected at random from amongst test-bank enrollees.)	HomeSense: Design of an ambient home health and wellness monitoring platform for older adults	This paper describes a pilot ambient home sensing project, HomeSense, actively deployed in the homes of older adults residing	While the number of sensing devices varies by the number of rooms in a home, most deployments utilize between 16 and 20 sensing devices to capture 90% home coverage.	A variety of methods have been suggested for elucidating the relationship between home-based behaviors and health computer vision and pattern recognition, application of fuzzy logic and ML use of ANN, BN , Poincaré Plots Markov Models, Ontology-Based Context Modeling and application of LeZi algorithms.

2.2. Reviewing Objectives of the Studies

After reviewing the goals of the articles and research and compiling Table 1.1., one would come to the following reasons, yet it should be mentioned that the ultimate goal of the

articles may be that it is necessary for the top of the correct order and instead of accurately distinguishing human behavior and the abnormal state from the normal one, it is necessary to understand and recognize the normal state of the house. one could consider the smart home management center(SHMC) as the human brain, just as the brain first examines the conditions and the environment before issuing any instructions for each part of the body, and then unlike normal states, the brain issues different orders and commands in abnormal situations. It is very important that the SHMC has the capability to detect the house condition and residents' behavior. As mentioned earlier, the main objectives of the previous studies could be grouped as follows:

a - In this new world, maintaining the security of material assets is perhaps one of the most challenging issues for human beings, and efforts to increase of the home security and its occupants are worthwhile. By recognizing normal behavior of residents in unusual situations, the level of home security can be increased, and could even be revalidated through the house occupants' type of behavior. For example, with a series of security questions to test if the person present in the house is a resident of the house or a stranger. In addition to protecting the residents' material assets, would bring about a greater sense of security for the residents of the house which is perhaps much more valuable.

b- Nowadays, due to the complexity of the smart home, even those people interested in technology have difficulty with the proper management of this home, and through recognizing people's behavior, some routine tasks can be assigned to the smart home management system, i.e. after ensuring the absence of abnormalities in the home. In normal cases one can predict future activities and request according to the status of sensors and past activities and hence assign certain tasks to home management system and give some commands without the need for direct instructions from the occupants. For instance, turning on the light automatically when the house is dark or turning on the air conditioning automatically when the indoor air needs to be changed. Such a routine not only makes the home environment enjoyable for residents, but also eliminates the possibility of forgetting essential tasks such as turning off some hazardous appliances. By reducing the control commands issued by the occupants of the house, it leads to the ease of controlling smart homes and thus increases its performance. But as mentioned before, the first principle and perhaps even the most important point of this case is to recognize the conditions and normal behavior of the residents of the house.

c - In order to extend the life of building equipment, majority of the task is delegated to the smart home management center(SHMC). Note that this equipment may be a simple electric switch or the entire building ventilation system. But for this, the first principle is that the SHMC must have the necessary capability to communicate with the home environment. For example, after recognizing residents' behavior and predictions, an electric switch may not need to be pressed several times during the day by residents or when occupants of the house are asleep, it is not necessary for the coffee machine or iron to be on. Moreover, all these managements will cause less damage and depreciation of home equipment and thus it is considered important to try to adapt the condition of the home equipment to the behavior of the residents in advance. In addition, residents' needs will most likely be met in the best way possible with the least amount of time utilizing the home equipment.

d - The number of single-person families has been increasing over the years around the world, and the majority of these families are elderly people, who do not want to go to a nursing home or do not have the possibility of having a 24-hour nurse. The capability to detect occupants' behavior and abnormalities in the home helps people live at a higher level of health in their homes. It goes without saying that this benefit is not just for the above-mentioned families. One of the simplest benefits of this is that in the event of an abnormality, the smart home management center can execute a series of predefined commands. For instance, it can inform the relatives who are close to the residents. Even at another glance, thanks to an intelligent management center, residents won't have a mental conflict about whether they have done some important tasks or not as well. Moreover, in some cases, such as the use of drugs for a sick resident, it may also be directly related to the his or her physical health. Therefore, part of the previous studies in this field is related to increasing the possibility of living only for the elderly and the sick by ensuring their health and maintaining it. With advances in this area, people may be able to automatically address treatment and care purposes for patients or people in need of help so that they do not need 24-hour care during their treatment (Deep et al. 2019).

e - In the introduction section, the issue of rising energy prices in general was addressed, saying that as energy values increased and energy savings went up, people became more inclined to their own smart homes which in turn made governments invest in smart homes. In this section, it is examined that by recognizing human behavior and home condition,

home management center can make decisions to better manage energy consumption according to the conditions. For example, when it knows that the office space will not be used in the next 24 hours, smart home management center turns off ventilation and lighting system of that room completely. Of course, in these studies, the goal is not only to reduce energy consumption, but on the one hand, it is to manage consumption, increase efficiency, and optimize energy consumption, and on the other hand, proper and efficient energy consumption to improve the quality of life. For instance, in winter, home management system can predict the condition of the house, which part of the house will be used at what time, and half an hour before the presence of the person to use the heating system of that room. Therefore, this optimization in all studies is not to reduce energy consumption, yet in some cases, it is to use energy properly.

2.3. Reviewing the Provided Methods

Due to the great importance of this field in smart homes, many studies have been done and considerable efforts have been made to enhance performance accuracy in all dimensions. One must look beyond study objectives and achievements to the main and computational part of the studies and with a more general look at the main part of the studies conducted in this master thesis, the proposed methods have been divided into three main approaches.

a - Based on the use of machine learning algorithms

b - Based on the use of deep learning algorithms

c - Completely unique modern methods based on mathematical and statistical calculations

All of the above methods have special features at a glance; however, if one aims to speak in general terms, it could be said that the first and second methods have received more attention than the third one. But this does not mean that one should ignore the third method. In some cases, a series of simple mathematical calculations have been much more useful than heavy calculations. As after the introduction of deep learning method algorithms and increasing the efficiency and accuracy of deep learning method compared to machine learning, attention to the use of deep learning method has increased in the past few years. However, this does not mean that the machine learning method is weaker than

the deep learning method, or the efficiency of deep learning methods has always been better.

Indeed, it is necessary to mention that in addition to the other three methods, some other methods may have been suggested, but due to their poor performance or limited interaction, their review has been ignored.

Therefore, in the implementation of studies, according to the purposes of their work, they have often presented their methods in one of the main ways. Also, it can be said that nowadays, due to the growth of computational accuracy and variety of algorithms in the first two methods, namely machine learning and deep learning, many studies have used those methods but again we should not neglect the third method. At times, presenting a series of simple mathematical calculations have been very practical and useful.

2.4. Datasets Classification

Apart from reviewing the proposed methods and the algorithms section of the studies, all the past works in terms of data can be checked. This means what datasets these studies have used to implement and test their method; moreover, how is these datasets gathered and how is it supposed to be obtained in a real smart home in the future.

The datasets that the proposed method of a study needs to be implemented is very important. Almost all previous studies have been reviewed and all the methods by which datasets were collected have been divided into the following five categories:

a - Using video and audio data

b - Utilizing sensor data such as motion sensors

c - Using simulated or uncollected data from reality

d - Using existing wearable devices data such as smartwatches

e - Utilizing more specialized wearable items separately with special design

To begin with, regarding the use of video and audio data, we can say that this method violates most people's privacy. In addition, it is possible that with the camera in all parts of the house, the feeling of comfort in it for the residents will be lost.

And in the case of data obtained through the simulator, one could say that the most important drawback of this information is the inconsistency with the facts. In some cases,

this data may not be true, and the methods evaluated via this data may not provide accurate and reliable results in the real world.

In the next step, the last two categories will be examined, and the fact that they have three important drawbacks and negative aspects will be pinpointed. The first disadvantage is that it should be carried by the residents all the time, and if forgotten by the occupant or the battery runs out of charge. Another downside is it may run out of battery power. Finally, forcing people to carry these devices might make them uncomfortable, so some residents will most likely refuse to carry these devices for this purpose.

The best way to use sensor data is known so far. Sensors are often very inexpensive, are fixed in the home, have a long life span, and their information is more accurate than all other types.

At the end of this section, if one wants to examine advantages and disadvantages of previous studies, it could be said that each study has its own advantages and is valuable. All these articles and studies have made the researchers reach a level of ability in this field nowadays, but if one looks at the negative aspects in more general terms, perhaps the most important weakness of these studies and articles can be considered as two cases, one in implementation section and the other in data set section, most of which are designed for normal people. Those people who have orderly and regular life and may not have acceptable results in recognizing the possibility of an abnormality in people who do not have an organized, normal life. And the issue of privacy in some approaches in some areas has been neglected a bit. In other words, in some methods, due to their approach, privacy has become a difficult issue. This invasion of privacy is possible both at the stage of recognizing the behavior of residents and at the stage of recognizing the abnormal state.

After reviewing all the available studies in the course, a series of overviews have been touched upon, which finally led to a series of classifications mentioned above, and now, with those attitudes, method of the study will be presented with more efficiency in the next sections.

3. HUMAN ACTIVITY RECOGNITION WITH AMBIENT SENSING DATA

3.1. Designing the Approach and Traditional Methods

In this section, the problem type will be subjected to scrutiny and the solution will be presented. As it was examined, the goal is very clear. One aims to be able to understand if there is an anomaly. And the process is first to predict the behavior of the person according to his or her circumstance, and then compare the predicted behavior with the behavior that the person has shown.

Finally, when a set of activities is performed by a resident of the house, one wants to find out whether that set of behavior is a set of normal behavior of that person according to all conditions and parameters or based on the personality and behavior of that person in the past, these behaviors of this person were not common and normal.

To better understand this approach, an example will be briefly reviewed. We consider a resident of a house. His or her behavior is tested to see whether it is normal or abnormal. For this study, one needs to have a model in order to compare the performed behavior, so a pattern of behavior from the behavior that a person has done in the past must be found; this model is made from a set of individual behavior in the past. In fact, we have an evaluation criterion with which to evaluate new behavior. This model, which is the behavioral model of each person, is an evaluation criterion and scale with which new behavior can be evaluated.

So the aim is to create a pattern or model for each person’s normal behavior. Creating a pattern would mean going through different paths and coming up with different ideas. Perhaps for each person, one can consider this model as a multi-unknown mathematical equation. In fact, this equation can take place with unknowns such as time, location, state of sensors, etc. After solving this equation, the output of it is determined as the normal behavior that a person should show at that time and place. But now the most important part of this story is to find the coefficients and the formation of this equation.

It is safe to say that in order to find an anomaly for any person, one must find a pattern in that person's past behavior. Nevertheless, a close look at the nature of human beings, one would realize that it must be really impossible to make this pattern, and if it is not

impossible, it would not be easy at all. So far it is concluded that the prevailing order and rule in that person's behavior have to be known.

In this way, the most difficult part of the job is to find the behavioral equation or the behavioral pattern of the residents. The below example aims to clarify the method and the overall strategy and a graph with a linear equation is considered to begin with. A mathematical equation for this shape will be written. After a little research and spending very little time, the equation can be found easily. After finding the equation of form, we can find ‘y’ in any expression with a few calculations. One can easily find out what the output should be.

If the human behavioral rhythm is like a graph with a linear equation, one can effortlessly form the behavioral equation of each person and be able to recognize normal behavior. In other words, when one encounters a number other than this number Y in X, he or she would say that this number is not real for the equation, is an unusual answer to it, and would also say a mistake has been made along the way and thus one finds the existence of abnormal behavior and state that an anomaly has occurred. So once one finds the real equation, he or she can easily detect it.

Now consider an almost simple diagram. One would make an attempt to form an equation for this graph as well, so that he or she can find the output quickly at any time without having to look at the figure. At first glance, one would understand that the following diagram was not simply the previous one, but again, the equation will be accessible by spending more time.

Now there is a quadratic equation diagram. For this graph, one also wants to form a mathematical equation to understand how the graph is formed. With a little effort one is able to discern that the diagram is not comparable to other shapes, and at a glance and with simple calculations an equation cannot be formed and it has many irregularities. Moreover, one may not be able to write a general equation for this graph.

The question which is posed here is what diagram human behavior would look like if it were a diagram; in addition, the answer would be that it would naturally be a few thousand times more irregular and lawless than any other diagrams. Now, assuming that human behavior can be implemented diagrammatically, the interesting point here is that those

who have not yet been able to form an equation for a slightly complex diagram, how can they form an equation for human behavior?

a mathematical equation is written for most graphs so when one comes across a very complex graph, he or she tries to find it by looking at the equation graph from above. In other words, one could make an effort to form an equation with a simple look at a complex subject. One way is to break the chart into small pieces, and write an equation for each part. By defining the numerical range and a series of methods, those equations could be assigned to the whole diagram and thus the conclusion can be reached which would be true at first glance. However, for the whole diagram an equation cannot be written at first, yet through spending time and changing attitudes one could write an equation for the small parts of the diagram. Now the material and this strategy are utilized in finding a pattern for human behavior. In order to understand the method more accurately, it will be explained in the most obvious way possible in the following.

It is known that naturally a few people have a very regular life, so that one can easily find that person's life pattern. Moreover, the problem is that even in the lives of regular people, there is no obvious order, and the fact is that human life cannot be seen as the life of a trivial robot. In order to find a pattern, one must act like a programmed robot in his or her life. For example, in the case of that person's sleeping habit, he or she must wake up at 8 o'clock every morning and go to bed at 11 o'clock at night. To say that it is unusual if a person goes to sleep at 9 o'clock p.m., i.e. earlier than usual, cannot be a right judgement. There is no such person in the world who has very regular and consistent behavior. Because part of a person's behavior depends on him or her, yet another part depends on the situation and environment and hence cannot always remain constant and predictable.

In the diagrams of the previous pages, equations were shown with one or two parameters but with a very simple examination; it was also mentioned that human behavior has hundreds of parameters and each of those parameters has a specific nature and a certain complexity, and it is much more complicated than those to be able to write an equation. In addition, by solving it, one can find out the type of behavior issued and say whether the behavior is normal or not. Therefore, one has to look at the problem very realistically and come up with a real solution.

So far it is somehow understood that there is no easy mathematical formula, and that real life is full of turmoil. Yet the golden point is that through study one could find that these disturbances are valuable, and from these disturbances life is born, while it arises from the order of habit. However, most people are not habit-oriented at all, and if they are so for a limited time, they will definitely break down the structure soon. Now the solution is the same structure-breaking people, i.e. it will be decided based on the amount of confusion.

To understand that a human being can be full of contradictions in his behavior, an example is given portraying that even in the case of highly disciplined people, it cannot be said that a person is always in order with a process throughout his life and owns the same set of behavior, ethics and tastes. For instance, he is very upset that his first child did not pass the exam at school and behaves in a wrong way, and may even blame his child, but the same person after years and gaining experience, at another time when his next child does not pass the same exam in the same situation is not upset with the same intensity and there is not even any behavior that is upsetting, such as punishing the child; hence, people may change their behavior by changing only one parameter, such as time. In addition to all this, it is obvious that human beings may even show different behavior in the same events, situations and time according to the people present at that moment. As a result, all parameters can have a direct impact on the outcome of the behavior.

So far the research has reached two very important conclusions. The first is that the most important need is to find the anomalies of the pattern of that person's behavior. The second case is considering that man is an autonomous being and involves many parameters in his decisions and choice of behavior, so he has no obvious and clear process and order, and thus one cannot practically find a mathematical pattern easily. As it was said in the first conclusion, the need to have a model for comparing people's behavior is crucial. Now the solution would be to take a look at those disorders and turmoil to find a way to the anomalies to be able to find an approach.

At first, it may seem that one can order behavior a little; in fact, the paper attempt to revamp the behavior from irregularities and confusions to the behavior with a pattern like Figure A which can at least be found by trying a pattern. For this transformation of disorder into order, there are two important aspects. The first aspect of this question is

whether it is possible to turn omelet into eggs, to turn a glass of cold water into pieces of ice or hot water, to turn pieces of firewood into a big tree. As logic says, these processes are not possible. That is why a disorder cannot be brought back to order. In other words, a disorder does not automatically become order.

This view stems from the problem of entropy, or the state of objects, and based on these laws, it is said that the entropy of irreversible processes never decreases with irregularity, but increases and remains constant at best. Therefore, if one generalizes these rules to the order of human behavior, he or she will find that energy is needed to have regular behavior and it is obvious that one cannot interfere in people's lives and behavior, so this attitude is put aside altogether and said that it is not possible for people to behave in a regular and orderly manner without spending energy unless they themselves want to do something regularly. Also, it will be very energy-consuming and time-consuming for them.

Once again, the research has reached a dead end in presenting the new approach, but with a change of attitude, one could move forward, i.e. one will have an overview of all behavior and events. To better understand, it is described that most events are disorder in themselves. They have, and at a glance, all events and happenings are chaotic, and in fact entropy dominates the natural processes of the world. Even some scientists believe that the entropy and irregularities in the universe are increasing. And as it has been said, according to a popular belief, entropy is called a state of disorder and unpredictability.

Now it is accepted that there is chaos or disorder in human behavior. But the important point here is that the irregularities are not just in human behavior. They can be seen in everything that happens on Earth, but where do you get the saying that the earth and galaxies are a regular set. In the continuation of this chapter, an example will be touched upon, and after that, a golden sentence will be expressed.

When one applies perfume in the corner of the room, its scent gradually spreads throughout the room. Thus, the perfume goes from order to disorder. But now the golden sentence is explained which will be the key, and from the next section according to this principle effort is made to find the anomaly. The golden sentence is that many scientists have accepted an order over all the irregularities that rule our universe. In fact, the world is in order at the height of disorder. To put the rest of the steps of this study in one sentence, one can contend that with this principle, the present research is trying to find

the existing order in the disordered behavior of human beings and following its own patterns through them.

So far it is understood it is true that human behavior often occurs without a specific order and pattern. However, from a higher perspective, some theorists contend that disorder in the scientific sense is a mathematical concept that may not be very accurately defined, but it can be considered a kind of coincidence with certainty. Its certainty is due to the fact that disorder has internal causes and does not occur because of external disorders, and it is accidental because disordered behavior is irregular and unpredictably accurate. As an example it could be said that perfume tends to be irregular when it spreads into the room, but with a closer look, one can find out that its diffusion is a principle showing how it is diffused and distributed with a general order and rule.

An example to illustrate this is dominoes. Instinctively, disorder evokes a picture of destruction and confusion, but one could find out that these irregularities are due to the complexity and inability to properly understand the relationships between the components of the disordered world. One might rightly imagine that every action and process creates the future. Every human thought and action, every leaf that falls to the ground, every water molecule that is constantly moving in the sea, contributes to the disorder.

In other words, every small cause can have a great effect and consequence. The theory is the familiar name of the butterfly effect in this case. According to the fractal model, the movement of the wings of a butterfly in the state of New Mexico can cause a storm in China! Although this process may take a long time, it would not be impossible. Similarly, when one sees Jehovah's storm, it is not without a reason. Disorder is not disorder but it moves according to its order, and one may not understand its order, or else one compares it with beings other than itself, and think it is disorder while it has its own order, and the same is true of human behavior. In other words, one may not be able to find order in a fleeting glance, but there are certainly patterns and templates in each person's behavior.

To reinforce the solution, one can say that, in the universe, a cause that is not the cause of anything can destroy the entire universe and lead to disorder. Whatever happens and behavior that follows is something else that depends on its normal necessity. Furthermore, whatever happens is looking for something that is usually related to it. So there is nothing that has come into being without a cause and a rule. If a movement intervenes in the

universe for no reason, then the world will be fragmented and lose its cohesion as a single system. Therefore, one can contend that humans perform their behavior in a specific and complex order, and examine other causes and relationships between events to discern human behavior.

In medical sciences, the opinion of the Greek philosopher has been confirmed, and even some scientists say that human existence is all in order. If man wants to study the biology of his being, he sees strange wonders in his being, and so is the order in the world of creation. The existence of order and rule in all the implementation of this world is not a subject that can be rejected and ignored, and in this regard, many scientific journals and books have been written that by studying them, the existing order in the world is known.

In spite of all the above statements, in the real world, one comes across behavior or events that are contrary to our perception and are similar to disorder. For example, a person with a healthy body suddenly becomes ill. In the past, having a sudden illness was the cause without a cause and without an order. But after the advancement of medical sciences and in more detailed studies, one could find out that the cause of the emerging disease in a healthy person is related to his or her nutrition, i.e. the food consumed by the infected person. One could simply get the idea that with the advancement of science and aristocracy to the rules governing the human body, the cause of the disease is also determined. Hence, one can contend that when seeing an event and feel it is out of order, it means that his or her knowledge and understanding are still incomplete and therefore he or she cannot detect its occurrence. In short, one do not accept disorder in cases which are thought to be disorderly, and one will move forward according to the philosophy saying that in the whole world, all behavior, events, happenings, etc. are regular with valid reasons and causes. Moreover, one would argue that all human behavior, at a glance and in general, have its own order and is subject to its reasons and causes, and is orderly in its own way. Once again, one could say that there is order in this world, and whatever we think is disordered has its own order indeed. Even if there is disorder, there is no problem. In fact, one does not need to find order and the pattern and cause of everything in the world in order to prove the order. Thus, the argument of order is in its own strength.

Among the discussions related to order and causation, two important points are: order does not mean restricting one's freedom, but it happens in order to rationalize one's

behavior and activities according to logical and specific criteria. The next and more important point is that this amount is order in everything that happens, so the search is to look for the wrong thing to do. It is said order does not mean that all behavior and events are in accordance with the previous plans; thus, one looks at those events and behavior, and use the points made in the past to try to find behavior which has happened abnormally. Of course, those two points will be explained in detail in the following sections.

So from a general point of view one do not accept the existence of a fundamental turmoil in the world and what happens in it, but rather all events happen for some reasons. Now, to generalize this theorem in human behavior, one could say that all these irregularities in human behavior have a unique order, rhythm, and pattern as well. For example, someone is very stingy. Although it is true that this morality is regarded as stingy, it may not be in all of one's behavior. However, by taking a comprehensive look, one would see that the person whose predominant behavior is based on being stingy and may behave abnormally in a wrong way to be generous for a moment, but does not know that that behavior is considered an abnormal one. As an example, one could refer to the miser story. One day a miser invites all his friends to lunch for no reason. As a result, people around us each have specific characteristics in every part of their lives. In fact, in a more comprehensive view, they have approaches and disciplines in their behavior. When it is said that a person is very studious, it means that most of the time he prefers reading and studying to having fun and resting. Yet another scenario is also possible; although that person has lessons he or she has not studied yet, he or she may waste his or her time playing and having fun.

Therefore, it is obvious that to detect an abnormal behavior, one must first build a pattern of behavior from the person's behaviors in the past. Of course one would say, this does not mean that one must use the existing order to find the anomaly after accepting the existence of order in human behavior, but rather one can look at things which can be determined by themselves through a series of standard calculations. In addition, one makes an attempt to find only an amount of these disturbances in order to use it to understand the existence of an anomalous state. To better understand the following sections, the issue will be discussed and forming a criterion and a model for evaluating a person's future behavior will be explained.

In this section, behavior which rarely happens will be discussed, and one would say that there is no need to put too much in the pattern for such behavior which rarely takes place, and spend one's energy on it. Thus, the pattern for such events will be followed. By changing a person's behavior, one could conclude that day shows abnormal behavior, so the abnormal state dominates his or her life and home.

After this point, one starts by paying attention to repetitive behavior. For example, when a person sleeps, it is not possible to say exactly when he or she does so, and when he or she wakes up. It must be said it is true that it is not possible to say the exact the time, but it is possible to recognize the sleeping process of that person, and according to that trend, if there is an abnormality in that person's sleep, the anomaly will be found. Hence, the existence of an anomaly cannot be found is a completely wrong sentence. Maybe with a few times the average and minimum and maximum findings one could find the exact anomaly. As an example, one could refer to a person who has always an average of 8 hours of sleep a night. Sleeping 3 hours a day indicates an abnormal behavior, or else a person who always sleeps at 11 o'clock at night at worst. Waking up at 2 in the morning is a sign of abnormality, so to find anomalies, instead of trying to limit patterns, one has to change his or her view from above. The most important point here is that one must not forget the principle that with an unusual behavior he or she cannot comment on the whole day of that person; thus, a set of behavior is considered in the view, and this look at a set of very personal reactions will make it much easier for us.

Therefore, it was concluded that there are some rules and order in all behavior, yet it is not very easy to find those rules, and they are seen as a series of disturbances in front of one's eyes, so one must use those disturbances. Moreover, to detect the presence of anomalies, one should first try to turn the amount of those perturbations into a series of equations. Additionally, by using those equations, one can find out how much disorder a person has in his or her life, i.e. he does not act according to his or her behavior in the past. It does not, so there is an abnormal state. Take a person who goes to the bathroom at least 2 times for a maximum of 2 times and for a maximum of 6 times during 24 hours, and stays there for 1 to 5 minutes each time he goes to the bathroom for example. If the range of these times changes drastically, one can say that an anomaly has occurred. for instance, that person has gone to the bathroom 9 times in the last 24 hours and the average

time he or she was in the bathroom was 11 minutes. Thus, one can guess that person's condition is not normal and he or she might become sick.

So far a number of results have been reached. Now a set of rules will be written in an attempt to find the patterns and templates, i.e. to find models and criteria in the behavior of homeowners. As it is known, perhaps the most important part of solving any problem is understanding the exact form of the question at hand. In fact, what question and request do we want to answer and the correct expression of the final goal is very important in achieving the desired results, so at the beginning of this section the desire will be explained clearly, and then a method will be provided.

In order to describe the ultimate goal in one sentence, one could say that in this master's thesis, a method is provided which can detect the residents' behavior on a dataset and the abnormality in the house. As it has been reviewed, previous articles and studies have had features, advantages, and disadvantages. In this method, it is aimed to find a way that preserves all the advantages of other methods, and at the same time tries to improve the drawbacks of those studies as much as possible. In this thesis, attempts have been made not to ignore the three principles in the proposed approach. In the following, those three principles will be described.

3.2. Principle One – Privacy of Residents

Perhaps the main feature of the thesis with previous studies in this field is the effort to fully respect the privacy of smart home residents. The privacy of individuals at all stages has been tried to be observed as much as possible. In the data selection section, a dataset was selected that did not use any wearables during the collection phase. Also, no camera or device for recording residents' voices has been used to collect data. The only data used was enough for the data sensors. Sensors which not only protect the privacy of individuals, but also have a long lifespan, low cost, and high variability with very high accuracy and for this reason, it can be said that in addition to trying to avoid the obvious violation of the privacy of the occupant of the house, perhaps the selected dataset, due to its nature, is one of the datasets that do not transmit any bad feelings to the residents (Ghahramani et al. 2018).

3.3. Principle Two - The Possibility of Severe and Daily Irregularities

In this master's thesis, an approach is introduced which is not in search of finding the order and life plan of every person. One can say that one of the most important features and characteristics of this research is that there is no need for a fixed plan in the lives of homeowners. Moreover, the ability to find the general pattern of life of the inhabitants of the house for any kinds of lifestyle. In other words, one aims to move forward in a way that is compatible with any rhythm of life. In fact, it should be responsive, just as it is responsive in homes where people live with a well-defined order and schedule, as well as in homes with residents where different events and happenings occur every day. Thus, one must not forget that there may be people whose every day is different from yesterday and tomorrow, so a solution has to be offered which works for everyone.

3.4. Principle Three - Using Light and Simple Calculations

The last principle to be considered is that offering a method requiring heavy processors is not aimed at here. In other words, in this way, one tries to pay attention to the fact that the proposed method does not require a heavy, advanced, computing system in terms of computing. In addition, it can be found for the management system in the smart home management system. Therefore, all these computational processes have been performed in the implementation phase of a system with intel CPU Core i3 8th-Gen and DDR4 RAM 4.00GB.

The three above-mentioned principles have been considered in all stages of the study, and it is very important to mention that almost all the articles related to the subject of this thesis have been reviewed. In addition, it is quite obvious that all the proposed methods have certain advantages over each other, but perhaps one can say that the biggest drawback of the previous articles is that while they are trying to find the anomalies of each person's life, these points have not been paid special attention. That privacy is very important, and it would not be possible to provide expensive processors for all smart homes in the next step. Furthermore, in most cases, many people do not have the same lifestyle, even those who have a steady government job. To better understand this point, an example of a nursing job in a hospital will be reviewed and for those people who do not have a fixed rule and norm, it can be said that the possibility of error is rather high.

To better grasp the idea, three examples will be given. Consider a university professor who lives a disciplined and routine life. A professor who visits the workplace on a regular basis and sleeps at night at the appointed time. When he does not show up for work one day, or when he is awake one night until morning, it means that an anomaly has occurred. But now consider the personal and professional life of a hospital nurse. At first glance, it can be said that he has a very simple pattern, he goes to work one day and rests the next, but on closer inspection, due to the nature of the occupation, this trend may change many times each month. In one week, due to the increase in the number of patients, the number of working days increases and in the next week, he may have to go to work for a few days due to the leave of one of his colleagues. It is seen that this trend may change many times. This was just an example of a person's work schedule. Of course, the rest of that person's life schedule will change due to this change in working hours. So in the life of a nurse, one may not be able to easily identify abnormalities in the life of a university professor.

A university professor’s not going to work on a working day may be an anomaly, but not all of it can be considered an anomaly to change a nurse's work schedule, so one needs to come up with a solution which can work well in all lifestyles. The significant fact is that there are many people like the mentioned nurse. As another example, consider a computer engineering student who sleeps a few nights a week at normal times, and stays up until morning to do homework for a few nights. Thus, a way has to provide in the next sections which can detect all these abnormal situations and behavior. In fact, it is an algorithm that is capable of detecting the abnormalities of the life process of severely disordered and irregular people. As it was said, if a professor does not go to work on a normal day, there is a possibility of anomaly, but it may be normal for a nurse to go to work on a day off. Therefore, one can witness how difficult it will be to find abnormal behavior.

As it was said in the previous paragraph, in this senior dissertation it is attempted to present an approach which is not based on a specific process for a particular type of person. In general, the harmony in people's lives at the height of disorder and apparent programming is aimed at. In this way, one does not want to have very complex and unusual calculations, but rather using statistical calculations and common machine learning methods is preferable. Now to put it in one sentence, decisions are made based on the number of changes in a person's behavior, and by looking at all these irregularities and complexities from above, one is trying to present a number that represents the number

of behavioral changes of a person, and it is for that person only. In the next step, to find the anomaly, it is decided, according to the amount of change in that figure per day, whether there was an anomaly on that day or not. Indeed, it is necessary to mention that in the proposed method, there is no attempt to find a specific pattern in a person's life, and attention is paid to the privacy of individuals.

3.5. Description of the Anomaly Detection Approach

Now one more explanation is provided, and an overview of the approach will be given in the next chapter. In order to diagnose anomaly, one should first try to predict the resident's behavior at home using statistical methods and machine learning. In the next step, by comparing the predicted behavior with the performed behavior, the situation of the house and the residents of the house will be attempted to diagnose. It goes without saying that a method based on a set of rules is proposed to detect anomaly.

In addition, in order to describe the approach to find an anomaly without preamble and without any explanation, one must first talk about a concept called NNC, which is a positive number. NNCs show the number of differences between the behavior performed by the resident and the predicted behavior for the resident. In the proposed approach of the computational part of the three NNCs with the names of main-NNC, daily-NNC and sub-NNC and the procedure is such that every day sub-NNCs are calculated according to the mathematical calculations considered inside the rules. And then sub-NNCs are added together to create daily-NNC. This daily-NNC is related to the behavior and actions of that person on that day. Moreover, after a few days, the main-NNC is calculated from the average of the previous daily-NNC of the last previous days. In order to know if the behavior of one day was unusual, one should compare that day's daily-NNC with main-NNC. If there is a huge difference, an anomaly can be suspected. The steps will be described in detail in the following sections. It should be noted that wherever 'NNC' is mentioned in this thesis, it means the main-NNC.

After explaining the previous concepts, ways of calculating sub-NNCs will be briefly discussed. It can also be said that in general sub-NNCs are calculated using functions calculated by statistical calculations and calculations based on machine learning algorithms. Therefore, there is a daily-NNC for each person every day, which is the sum of sub-NNS.

As mentioned above, there are two types of calculations in sub-NNCs. The first group is based on statistical calculations, and the second group is based on the use of machine learning algorithm. It is very important to mention that in the next steps it will be examined whether the accuracy of the method is increased by using deep learning; in other words, sub-NNCs are designed in a way to have rules inside benefiting from deep learning, i.e. instead of using machine learning, deep learning is used.

4. EXPERIMENTS AND RESULTS FOR HUMAN ACTIVITY RECOGNITION

4.1. Designing the Approach

After reviewing the initial points in previous chapters, such as the reasons for the importance of this issue, past studies, and description of the issue as well as providing an introduction to the approach and setting out a set of goals, it was decided that the method should include all the goals mentioned in the previous chapters. Moreover, it should have an acceptable efficiency and productivity. In this chapter, all the requirements for understanding the proposed method will be explained. Moreover, providing a way to detect abnormalities in smart environments is of utmost importance.

Although it is necessary to have all the previous sections, but this chapter is perhaps the most important part of this master's thesis, so an attempt should be made to describe everything carefully. In this chapter, being ‘rule-base’ will be defined first. Actually, it will be explained what it means to be a ‘rule-base.’ Then, the concept of sub-NNC will be explained. After that, the two concepts of main-NNC and daily-NNC will be discussed, and it will be attempted to achieve the main-NNC for each person by using Daily-NNC. Finally, ways of diagnosing anomalies will be touched upon. Of course, in the meantime, and after completing the whole method, one realizes that an efficient machine learning algorithm in the sub-NNCs based on machine learning is needed. Therefore, one has to look at the available algorithms in machine learning and from those algorithms, the best one has to be chosen; As it was said, choosing which machine learning method to use is itself an important issue because it will have a direct impact on the performance of the proposed method, so it needs to be studied and evaluated very carefully.

One of the most important parts mentioned in this chapter is that it is necessary to have a dataset to implement, test and evaluate all the steps. This means that the datasets related to smart homes should be checked, and the best dataset should be selected; moreover, after selecting the dataset, the reasons for choosing it will be discussed, and the selected one will be examined and analyzed.

Now, after getting acquainted with the basic concepts required and selecting the dataset, one should refer to sub-NNCs which have rules based on machine learning techniques. In this part, an optimal and accurate algorithm of machine learning will be used. In the

first step, all the common algorithms will be examined, and for more detailed studies, a number of algorithms suitable for the method and dataset will be selected. In the next step, the performance of the selected algorithms on the selected dataset will be evaluated, and after receiving outputs of the evaluations, an algorithm with the best performance will be selected.

After all the aforementioned steps, ways of recognizing the abnormality of a resident's behavior is examined. This means that in the end, daily-NNC and main-NNC will be calculated, and then by comparing daily-NNC with main-NNC, the existence of abnormal states and behavior will be judged. In other words, the result of comparing two numbers will be a criterion for this study, and if there is a possibility of an anomalous state, it will be determined.

4.2. Description of the Rule-Based Concept

The whole proposed approach was touched upon previously, and as it was said in the previous sections that perhaps one of the most important differences between this method and the one that has already been proposed for finding anomalies is that it is rule-based. In this master's thesis, different explanations of the proposed method have been presented, and the fact of using the prepared written rules to diagnose anomalies was encountered several times. Now, in an attempt to give a very basic definition, one should say that being rule-based means that there are a series of computational commands and rules which examine the information obtained from smart homes to decide whether there is an anomaly or not. Actually with these rules, it is attempted to compare the degree of adaptation and similarity of the behavior performed by the residents of the house with their previous behavior.

Although the rule-based concept has become more tangible for the readers so far, it will be touched upon to an acceptable extent in this chapter. It should be said that the most obvious feature of the superiority of the rule-based method over previous methods is the possibility of its simple expansion and improvement. In fact, the proposed approach does not use a method limited to a specific algorithm to solve this problem.

This approach is not limited to one unchangeable principle in any part, neither in the data part, nor in the number of inhabitants and in the part of the algorithms used in the rules.

Take the rules using machine learning techniques for instance. If a new and more efficient algorithm is introduced in the future, the previous algorithm will be replaced with a new one. However, in this direction and in the implementation parts, one has to follow some principles of object-oriented programming so that no problems will arise in the future if one needs to change the algorithms with higher performance. Of course, it is not only the replacement of a new algorithm with the old one. Even with the emergence of a new algorithm or method, a new rule can be defined and that method can be utilized along with the previous rules to find anomalies. Thus, in this approach, all the sections with more up-to-date methods, data and sensors can be improved and upgraded.

As a basic definition, it was said that rules are a set of computational commands that when smart home data is given to these calculations, compares the behavior issued by the resident with his or her behavior in the past after a certain path, and concludes whether according to the behavioral pattern of that person, he or she was behaving in a certain way at that moment and situation or not. An example to better understand the concept of rules could be that a rule to detect anomalies can be based on the number of repetitions of each activity.

For example, in Figure 4.1. a smart TV can be seen having the natural capability to announce the number of times it is turned on and off. The state of the house and the normality of residents' behavior with a definition of a rule for turning the TV on and off is attempted to be understood. In other words, without the need for any additional sensors and costs, and sophisticated technology, and with the people's privacy no matter what TV show they watch. A rule is defined which finds the normal number of a particular behavior for a person, such as turning on the TV during the day, and whether the number of times he turns on the TV exceeds the normal number, through which a change in behavior is determined. Indeed, the amount of change in behavior is related to the amount of difference between the normal number and the performed one.

To better understand, the previous example will be explained further. If the resident has turned on the TV 2 times on average in the past, and turns it on 4 times on a new day, it shows that there is a difference, and should be specified; however, if he or she turns it on 8 times, the amount of changes is very drastic and thus the magnitude of the changes should not be ignored.



Figure 4.1. Photos from inside a smart home (URL-1)

Therefore, this rule can calculate the average, i.e. maximum and minimum numbers of turning on the TV, or it can even calculate the average usage time each time the TV is turned on and hence different calculations can be inside the rule. In the end, that rule is used by comparing the results of those calculations with the data of the new day to reach more accurate conclusions about the abnormality of the resident's behavior of the house. As it turns out, this rule is not only useful for turning on the TV, but one can also use the same rule for the number of times a person goes to the terrace, bathroom, or kitchen. As it is seen, with one rule one can find lots of abnormalities. In fact, a person who goes to the bathroom 10 times a day may be ill, or multiple visits to the terrace or kitchen may indicate an abnormal condition.

Thus, it is attempted to provide a method with perhaps very simple and fast combinations of available methods which are based on a series of predetermined computational rules which are like a path. This path is like a metric and a means of evaluating the norm of residents' behavior. After performing this set of calculations along the path, the residents' behavior should be assessed to be able to compare it with their behavioral pattern in the past.

In this thesis, a fixed path instead of a fixed algorithm is used to detect abnormal behavior of residents which leads to finding a pattern of behavior of each resident according to the

conditions of the house. This means that the data obtained from the set of behavior of the resident of the house is evaluated after a computational path such as finding the means, maximums and minimums, etc., along with a machine learning algorithm. In the first step, according to the calculations performed, and the results obtained from the comparison of the set of resident's behavior during the day, his or her behavior receives consideration. Moreover, in the next step, if he behaved abnormally, how much of his behavior were unusual. Therefore, the very important point is that with a behavioral change, it is not quickly concluded that a resident is abnormal, but it is decided with a set of behavior and with a look from above which is perhaps one of the reasons for the increase in accuracy (Zhou et al. 2015).

Since important features of the proposed method will be discussed at the end of this chapter, two important points should be observed in the process of preparing the rules; those ones which can be upgraded and changed in the future, especially those rules which are based on machine learning. The emphasis is on the fact that in some cases with very simple changes one may reach different results. The next point is that the approach is not attempted to judge too soon. Thus, it starts with a small change in behavior; it is not immediately concluded that the person's behavior was abnormal yesterday, and he or she is behaving abnormally today, yet it is said that this kind of behavior corresponds to a certain degree different from the past behavior, and the main anomaly in the set of resident's behavior is looked for and is decided with a set of his or her behavior. It should always be kept in mind how much the dweller has behaved erratically and in an unplanned manner in the past few days. In other words, the behavioral pattern has been abnormal and with that piece of information, one could give the possibility of unusual conditions for that person.

4.3. Description of Sub-NNC

In the previous sections, the rule-based concept was touched upon, the reason behind deciding to write a series of rules to solve the problem was mentioned, and said that small anomalies are attempted to be found with the help of these rules which are able to change behavior. The resident's behavior will be examined, i.e. one aims to see how different the resident's behavior today is from his or her behavioral pattern in the past.

After getting familiar with the approach and the rules in this chapter, the path will be subjected to scrutiny a little more clearly and ways of predicting each person's normal behavior in normal conditions will be explained, and the abnormalities will be diagnosed more clearly. In order to so, it would be necessary to put them in containers and pack them, i.e. to make a module. Thus, in this approach, after the concept of rules, sub-NNC will be introduced, and then it could be said that sub-NNC implement written rules as separate modules. After calling and performing a series of predetermined calculations, it would give a non-negative number in the output which indicates the number of non-routine resident behavior. To be more precise, the output number indicates the intensity and extent of changes in the resident's behavior from his or her behavior in the past, and expresses the extent of the person's unusual behavior. Therefore, the output of any sub-NNC is a large number indicating that the person has behaved outside of the irregular behavior, and whenever the number tends to zero, it means the person performs his or her previous fixed behavior.

For example, in a sub-NNC the rule examines the number of times a task is performed per day. If the output number of sub-NNC is small, it means that the person has performed the same number of times as before. For example, brushing activity and the number zero means that the person has brushed the same number of times as in the past, and no special changes have taken place. Therefore, from now on, one knows the rules in the form of sub-NNCs. Moreover, sub-NNCs are used to find out how much has changed in the resident's behavior compared to his or her past behavioral pattern.

Sub-NNC uses basic calculations created by rules to show how much behavior they have been performed outside of the previous resident's pattern, and as it was said, every sub-NNC has only one output which is a non-negative number. Therefore, when each sub-NNC calculation is completed, a non-negative number is given in the output which in the future indicates the presence or absence of an abnormal state in the behavior of the resident of the house. In very simple terms, one can say that this number shows how much change in the resident's behavior has been found, and if this amount of change increases sharply, one will conclude that a large number of the resident's behavior have changed, and thus the set of resident's behavior has become abnormal.

It was said that every sub-NNC gives us a non-negative output after the calculations. On the other hand, we have several different sub-NNCs which are written based on different rules. Consequently, several positive numbers will be obtained after running all of those sub-NNCs. In the next part, the discussion will be how those numbers help one identify any abnormalities in the resident's behavior.

4.4. Definition of Daily-NNC and Main-NNC

Perhaps the most important concept in this study is main-NNC, so one needs to get acquainted with it first. In order to describe main-NNC, one must know daily-NNC and learn how to calculate it, i.e. to know where it is obtained from. Therefore, one almost has to start step by step from sub-NNCs, and after fully understanding daily-NNC, the definition and ways of calculating main-NNC will be touched upon. As it was mentioned, there are a number of sub-NNCs which after their implement, each of them would present a positive number in the output.

Now the first step after performing these sub-NNCs is to specify the task of these numbers. There is the easiest way to do this because in addition to its simplicity, the reviews have shown that it also has the best performance. All the outputs of sub-NNCs are added together and the sum of sub-NNCs of that day constitutes the daily-NNC of the same day, i.e. all the written rules of the results are added to the daily-NNC. The Daily-NNC shows how much behavior the resident performed that day which did not conform to his or her own pattern of behavior. In very clear terms, Daily-NNC is a parameter that shows to what extent has the resident performed his or her behavior in accordance with the predictions made by computational rules.

As it can be guessed, the importance of rules and of course related sub-NNCs in identifying abnormalities in behavior can be different. Thus, if the outputs add up simply, daily-NNC may not be able to help one accurately diagnose abnormalities. To better understand this, one could say there are two changes in the resident's behavior. The first is that the he or she has read a book twice as long as normal days, and that the resident has been asleep twice as normal days. It is quite obvious that he or she does not equally depict the existence of abnormal behavior. It is true that a long study may indicate that the person is bored or that there is an exam, and that there may be unusual circumstances, but it is very clear that if a person who usually sleeps 6 hours a day, sleeps 12 hours one

day, it is much more likely to have an abnormal condition, or maybe he or she has become ill.

To talk about the important difference in the type of behavior and the existence of differences in them, three different scenarios could be presented. The first is not to pay attention to it at all, and the next method is that output of each sub-NNC has a coefficient and with the help of those coefficients one can increase or decrease the importance of behavior. The magnitude of those coefficients determines the importance of the output of each sub-NNC. The last method is to apply those coefficients inside sub-NNCs, i.e. inside the rules to be calculated according to the nature of the output behavior of sub-NNC. If there is a severe anomaly in the behavior of the resident, sub-NNC gives a large positive number in the output. This method is smarter than the two previous methods, and in addition to solving the problem, it will not affect the way daily-NNC is calculated. Hence, it is calculated as before and only with a few additions of positive numbers.

Finally, daily-NNC is calculated and after finding it and over time, daily-NNCs form the main-NNC, so with a daily-NNC one cannot calculate the main-NNC. As a result, several days are needed after the start to form daily-NNCs. The minimum number of days is selected to form the main-NNC according to the size of the dataset. If the dataset has thirty days of information, the first ten days are enough to form the main-NNC, so far it can be guessed that the main-NNC is the average of daily-NNCs.

Then, the main-NNC is calculated by taking the average of daily-NNCs of the past days. Therefore, in the end and after a few days, there will be one main-NNC for each resident of the house and one daily-NNC every day. Now to diagnose an anomaly, one could compare this daily-NNC number with the main-NNC. In other words, if there is a huge difference between these two numbers, one can understand the possibility of existence of abnormal conditions in the dweller's behavior.

To explain further, it can be said that when the information of a new day is obtained, it should be checked whether there was an abnormal behavior of the resident of the house or not. First, with the obtained data, sub-NNC calculations are performed and daily-NNC for the new day is found. After calculating daily-NNC, one could examine how different daily-NNC is from the main-NNC. In fact, one can look at how much the amount of abnormality in the resident's behavior has changed on the new day compared to the

previous days. At a glance, it could be said that the main-NNC shows the extent of irregularities in behavior of the dweller. As it was mentioned, sub-NNCs are utilized to identify irregularities and unexpected behavior in the resident's behavior. Daily-NNC is the amount of irregularities in the behavior the resident has performed on the new day. Therefore, the existence of a sharp difference between these two numbers can indicate that the behavior of the resident of the house has changed, and there is a possibility of anomalies in his or her behavior.

It was mentioned that the two numbers of daily-NNC are compared with the main-NNC, and if there is a difference between these two numbers, one will realize the possibility of anomalies in the behavior of the resident of the house. Now, one point is how much change can be the criterion to have an anomaly state. Strictly speaking, how much daily-NNC of the main-NNC must be large or small to conclude that there is an anomaly. A fixed number is not specified for this, so to increase the accuracy of the anomaly diagnosis, the highest percentage of normal changes is specified, and changing more than that percentage, will help detect the occurrence of the anomaly. This percentage is due to the amount of change in daily-NNC on normal days in the past. For example, according to the amplitude of fluctuations of daily-NNCs in the past, if more than 30% of daily-NNC change compared to the main-NNC, there is a possibility of anomaly, and if the change is more than 50%, the anomaly happens with a very high probability.

Up to this point, almost all the important concepts have been touched upon, and ways of diagnosing the anomaly have been delineated. As it was mentioned, it is one of the most important and main concepts of daily-NNC. In order to calculate daily-NNC, there is a concept called sub-NNC which should not be forgotten, i.e. these sub-NNCs are aggregated after calculation and form daily-NNC. They will be explained in a little more detail. Moreover, a general grouping will be provided based on the type of calculation which is made within them.

It was stated that sub-NNCs are grouped based on the calculations which are done within them. This means that they are grouped based on the type of rules inside them. In all these sub-NNCs, the rules are mainly based on two types of calculations. The first group performs calculations based on statistical techniques such as finding the mean, the maximum and the minimum, etc. Perhaps this group of sub-NNCs may not be so critical

for detecting anomalies at first glance. But a closer look reveals that these calculations are very effective. Several examples have been given in previous sections to prove how important and vital this group of sub-NNCs is. But again, a very simple example will be given that by finding the average sleep time and sleep time range of the resident in the house— if these sleep hours change drastically— the possibility of an anomaly will be easily guessed. This means that if a dweller who usually sleeps from 10 p.m. to 6 a.m. and the average time spent in bed is 8 hours, sleeps from 11 p.m. to 11 a.m. (12 hours), there is a high probability of abnormality.

The second group of sub-NNCs will be computational based on machine learning techniques, i.e. the rules within this group originate from machine-based methods. It is important to mention that in the following parts a machine learning algorithm will be chosen according to its efficiency and accuracy, and will be used for training and predicting inside sub-NNCs on the dataset obtained from the smart home. Thus, there are two sub-NNC groups in the first group which is related to statistical calculations. The three sub-NNCs are described in Table 4.1.

Table 4.1. Description of sub-NNCs based on statistics

	Based on Statistics
S1	For each activity at each hour, the minimum and maximum of continuous spent time during the last 15 days is extracted. It is compared with test days' results; if for each hour of an activity she/he spends more than the maximum, sub-NNC wil increase by one unit and also if the spent time for activities is less than the minimum, sub-NNC will increase by one unit.
S2	For each activity at each hour of the day, it was ranked from 0 to 15 based on that activity's frequency at that specific hour in the last 15 days. Each activity at each hour of the test day is compared with the above-mentioned rank list; if an activity at a specific hour is done on the test day which has a frequency less than three times, sub-NNC will increase by 1 unit, and if an activity which has a frequency more than 11 times in the rank list, and is not done in the test day, sub-NNC will increase by one unit.
S3	Sub-NNC is computed utilizing the following three components: i- For each of the 27 activities, the sum of spend time is calculated, then is compared with the average of those activities in the last 15 days. If the variation of the sum of activities is more than 50% of the average one, then sub-NNC will increase by one unit. ii- For activities done on a test day, if the frequency of activities is less than three times in the last 15 days, sub-NNC will increase by one unit; meanwhile, for activities which are not done on the test day, if the frequency is more prominent than 12 times in the last 15 days, sub-NNC will increase by one unit. iii- Considering the last 15 days, the daily sum is computed for each activity; then, the maximum and minimum spent time for each activity in the last 15 days is extracted. If on the test day, he/she spends less than the minimum, sub-NNC will increase by one unit, and if the spent time is higher than the maximum, sub-NNC will increase by one unit.

Two of machine learning-based sub-NNCs were described in the Table 4.2. However, it is recommending to refer to the codes for a more practical understanding.

Table 4.2. Description of sub-NNCs that are based on machine learning

	Based on Machine Learning
M1	7200 seconds from the test day were selected randomly, and then the prediction is generated for activities for each 5% of error in prediction. sub-NNC increases by one unit.
M2	With the trained model, for each hour of the test day, the activities are predicted (If an activity lasts for more than a minute in an hour, that activity is considered to be done), and are compared with the actual label of activities in that hour for each activity prediction error. sub-NNC increases by one unit.

4.5. Sub-NNC Based on Statistical Techniques

In this category of sub-NNCs, the main calculations are based on techniques and methods available in statistics, i.e. the rules that exist in these sub-NNCs try to identify and reveal abnormal changes in the resident's behavior by performing basic calculations such as finding the maximum and minimum and the average. At first glance, trying to find anomalies with the help of statistical calculations may seem like a trivial task, but finding the maximum and minimum in an activity over the past few days has been so effective in some cases that complex algorithms may not be as complicated.

To understand the above claim, two very simple examples will be given that can help clarify it clearly. For example, the average number of times behavior is performed can be found. In fact, if a person who goes to the bathroom for an average of 5 times during the past days, goes to the bathroom 16 times a day instead, this can be regarded as an abnormal behavior. To make sure that this behavior is abnormal, one can also refer to the maximum and minimum number of times he or she has used the toilet in the past few days. On the other hand, if the resident goes to the bathroom the same number of times as before with the difference that he or she used to stay in the bathroom for an average of 4 minutes, but he or she stayed there for an average of 17 minutes on the new day, it could be said that the resident behaved abnormally and this behavior may indicate his or her illness.

The next example could be that the average can be used to determine a time frame for the behavior of the house resident. To better understand and conceptualize this example, one can consider a resident of the house who in his or her regular days, on average, usually sleeps 8 hours a night and sleeps between 11:30 p.m. and 5:15 a.m. This resident always gets out of bed before 8 o'clock in the morning. But on a normal, non-holiday day, there is a possibility that the same person goes to bed at 9 o'clock at night and sleeps until 11 o'clock in the morning. From these changes in the sleep pattern and discipline of the resident, the possibility of abnormal conditions arises.

It was shown with the simplest statistical calculations that small anomalies can be found. As another example, consider a researcher who, after statistical analysis, finds that the dweller usually sleeps only at night. Thus, this resident never sleeps after lunch, yet one day, he or she sleeps for three hours after lunch. By comparing this behavior with the person's behavioral pattern in the past, one could find that sleeping at that time is unusual for the resident.

At the end of the section on Sub-NNCs based on statistical techniques, four important points will be pinpointed:

a - The first and perhaps the most important point is that a set of calculations is performed in sub-NNCs for each individual’s behavior. In other words, all residents’ behavior is compared with their behavior in the past, and it is not that only certain behavior is selected, that behavior is evaluated, and after these reviews, any abnormal changes in each behavior is identified.

b - The second point is that the day is divided into 24 parts, i.e. it is examined hour by hour, and for each hour it is checked what behavior is likely to occur. For example, at 11 o'clock in the morning, there is no possibility for the resident to sleep or take a shower, because it has been found that at 11 o'clock, he is not often present at home. Therefore, in addition to looking at the type of behavior performed by the resident, one also has to check with the time of occurrence, i.e. what behavior usually happened in the past during the specified time period.

c - The third point is that all behavior is dealt with an overview and making hasty decisions is prevented regarding abnormal behavior, although some behavior may be

more important, such as not sleeping for the entire 24 hours or not going to the bathroom in the last 24 hours. Consequently, in this method, a set of behavior performed by the resident is taken into consideration. According to his or her set of behavior, one can say whether the resident's behavior is abnormal or not. Indeed, one can have an error or alarm for any anomaly which occurs, so that he or she can be informed of the anomaly on a case-by-case basis.

d - The fourth and final point is that in sub-NNCs rules are implemented using statistical calculations. A lot of attention is paid to the minimum and maximum, and in some cases where one cannot comment definitively, the possibility of an anomaly could be guessed with the help of those sub-NNCs. In fact, if there is a possibility of an anomaly, it will inform and give one a warning, and indicate that there may be abnormal conditions.

4.6. Sub-NNC Based on Machine Learning Techniques

Sub-NNCs were categorized into two groups, and now, after examining sub-NNCs based on statistical calculations, the second group of sub-NNCs, which calculate internal rules based on existing machine learning techniques, will be touched upon. Within this group of sub-NNCs, there are more complex and heavier calculations than the previous one, so sub-NNCs based on learning techniques need more time to perform calculations than the previous group of sub-NNCs. Therefore, smart home management center is more involved in computing and processing. However, in the face of this computational complexity and time-consuming nature, these sub-NNCs have the capability to predict the resident's behavior of the home in a variety of situations.

The fact that this group of sub-NNCs are capable of predicting the resident's behavior in more diverse conditions means that through using machine learning algorithms for each resident using his or her behavior data in previous days a pattern is formed for his or her normal behavior. In other words, they do not proceed only on the basis of statistical calculations, but the machine is trained with the data of the previous days to build a special machine for each resident of the house. This machine also pays attention to the conditions of the house while performing the behavior, so that the most important difference of this group of sub-NNCs is their attention to the state of sensors. More precisely, in this group

of sub-NNCs, the machine is trained in the beginning to be able to predict its behavior at any time based on the state of sensors and the time of events.

Thus, in statistical calculus sub-NNCs, one only looks at the data received from the past, but in machine learning sub-NNCs, one looks at the present as well as the past, and always pays attention to the conditions of the house and the dweller. Then, without knowing the status and data obtained from the home at the time of the behavior, one cannot decide on the type of behavior that may occur from the resident.

In order to talk about the general steps of this group of sub-NNCs, one can refer to the fact that these sub-NNCs build a machine of the resident's behavior first, and then the machine predicts normal resident's behavior according to the conditions of the house and the dweller. Finally, the predicted behavior is compared in different ways to the behavior performed by the resident, and any differences in the predicted behavior with the performed behavior are the criteria to decide whether there is an anomaly. Thus, the existence of an anomaly is guessed based on those existing differences.

Within Sub-NNCs, one does not move forward with a consistent general policy, and does not make hasty decisions about the resident's behavior. To better understand, if there is an abnormality in the behavior for a few seconds, it is no concluded that the anomaly has occurred. For example, at three o'clock in the morning, all the sensors are in a position indicating that the resident should be asleep, but the behavior of the resident is other than sleeping activity. In this case, one does not immediately say that an anomaly has occurred, but to make sure, the time interval is first changed to a few minutes. Then, the resident's behavior and data received from the house is re-examined. The reason behind this is that the resident may have woken up for a few minutes to go to the bathroom or he or she has become thirsty and has gone to the kitchen to drink water. Therefore, the fact that he or she is not asleep at that time is not considered an anomaly, since it is normal for a resident to be thirsty or need to go to a toilet during the night. But when the number of times going to the kitchen or bathroom increases or the length of time a resident stays in the bathroom or kitchen increases, the possibility of an anomaly arises, but one must also refer to past resident's behavior for a more definitive view. A person's past should also be taken into account because it is possible that the person has a permanent illness, and has to go to the bathroom frequently all nights.

The fact that a person's past behavior should also be considered does not mean that a new category of sub-NNCs is defined, since in each of the two sub-NNCs attention is paid to the behavioral data of the resident of the house in the past. The trained machine of a resident who sleeps in the afternoon is completely different from a machine of a resident who does not sleep in the afternoon, and naturally these differences originate from the residents' behavior of the house in the past. Therefore, machines consider the resident's past behavior before predicting his or her normal behavior (Dilraj et al. 2019).

In all the above sections, it was said that this group of sub-NNCs is based on machine learning, so one must determine which algorithm and method of machine learning is used. In order to find the best way to train the machine, in the next section several common algorithms will be discussed, and after comparing the results of the experiments, an algorithm will be chosen, and machine learning algorithm will be used inside sub-NNCs. The very important point of this section is that the accuracy of the selected algorithm has a direct impact on the accuracy of anomaly detection by smart home management center. This means smart home management center will find anomalies with the help of those two groups of sub-NNCs and the rest of the process. The accuracy of the previous group of sub-NNCs cannot be accurately assessed, but to trust the efficiency of the method, it is very important to choose a machine learning algorithm which provides acceptable accuracy in predicting the resident's behavior of the house.

4.7. Experimental Setup for Investigating the Performance

According to the above description of Sub-NNCs, a machine learning algorithm in implementation needs to be used in sub-NNCs. In this section, one of the algorithms can be randomly selected and used, but due to the importance of this group of sub-NNCs, one should be very careful in choosing the machine learning algorithm. Different reasons could exist behind it, but the important one for this sensitivity deciding which machine learning technique to use can be directly related to the performance accuracy of the proposed method with the accuracy of the selected algorithm. With this algorithm, the machine is trained to predict the residents' behavior of the house which means that in order to increase the accuracy of the proposed method, a more accurate algorithm must be chosen.

Accuracy plays a very important role, but one must be careful not to choose algorithms which cannot be implemented, and in choosing the two algorithms the number of calculations and the amount of time it takes the processor to do the calculation should be taken into account. Therefore, in two algorithms with a certain amount of accuracy the number of calculations might be different, and one algorithm could impose more computational load on the CPU of the smart home management center. It is very likely to have two algorithms with the same degree of accuracy one of which needs more time to find the result, or requires more computational resources, i.e. an algorithm with 4 GB of RAM can be calculated, but an algorithm with the least possible capacity will require 16 GB of RAM. As a result, one should be very alert to find the best algorithm for sub-NNCs based on machine learning.

As it was mentioned, the accuracy of this study's approach depends on the selected algorithm. Thus, in this section it is attempted to have an efficient algorithm and choose the best machine learning algorithm available now. To find and select the best one, the four main steps will be defined in the following order:

a - The first step is to pre-select the available algorithms for review. Those algorithms capable of predicting the residents' behavior based on the data are selected. Therefore, algorithms are looked at based on the data available from the home environment in the dataset, which means that in general one algorithm may have advantages over another one. However, due to the nature of the dataset, implementing would be impossible.

b - In this master's thesis, the second step will be reached after going through the first step, and the algorithms which can be used on the dataset in the future will be selected. In this step, a series of criteria for selecting a better algorithm will be expressed, the decision will be made according to these criteria. Next, they will be described and their definitions will be given. Meanwhile, those four parameters will be touched upon first.

4.8. Performance Metrics

As seen in the Table 4.3., true positives and true negatives are the observations which are correctly predicted and therefore shown in green. In an attempt to minimize false positives and false negatives, they are shown in red (URL-2).

Table 4.3. Classification of predictions (URL-2)

	Predicted Class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

True Positives (TP): These are positive values which are correctly predicted. It means that the value of actual class is yes and the value of predicted class is yes as well. For instance, if the actual class value shows that the passenger survived and the predicted class tells you the same thing.

True Negatives (TN): These are negative values which are correctly predicted. It means that the value of actual class is no and the value of predicted class is also no. for example, if the actual class shows that the passenger did not survive and the predicted class tells you the same thing.

False positives and false negatives happen when the actual class contradicts with the predicted class.

False Positives (FP): When the actual class is no and the predicted class is yes. For example, when the actual class indicates that the passenger did not survive, but the predicted class tells you that he or she will survive.

False Negatives (FN): When the actual class is yes, but the predicted class in no. For instance, if the actual class value shows that the passenger survived and the predicted class tells you he or she will die.

Once those 4 parameters are comprehended, Accuracy, Precision, Recall and F1 score can be computed.

Accuracy: accuracy is the most intuitive performance measure and is simply a ratio of correctly predicted observation to the whole observations. One may suppose that having high accuracy would be the best model. It is true that accuracy is a great measure; however, only when there are symmetric datasets, values of false positives and false negatives are almost the same. Therefore, other parameters have to be looked upon to

evaluate the performance of the model. For that model, the figure is 0.803 which means it is approx. 80% accurate.

Accuracy = TP + TN/TP + FP + FN + TN

Precision: it is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric aims to answer is: ‘of all the passengers labeled as survived, how many actually survived?’ High precision relates to the low false positive rate. It has 0.788 precision which is seemingly good.

Precision = TP/TP + FP

Recall (Sensitivity**):** recall is the ratio of correctly predicted positive observations to the whole observations in the actual class - yes. The question ‘recall’ answers is: ‘of all the passengers who really survived, how many did one label?’ ‘recall’ figure is 0.631 which seems to be fine for this model as opposed to 0.5 above.

Recall = TP/TP + FN

F1 score: it is the weighted average of ‘precision’ and ‘recall.’ Therefore, that score takes both false positives and false negatives into account. Intuitively, that is not as simple to realize as accuracy; however, ‘F1’ is usually more useful than ‘accuracy,’ especially if there is an uneven class distribution. ‘Accuracy’ acts best if false positives and false negatives cost the same. If the cost of false positives and false negatives are very diverse, it would be better to study both ‘precision’ and ‘recall.’ In that case, F1 score is 0.701.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

4.9. Performance Results

c - In the third step, one must have an evaluation of all algorithms under the same conditions, i.e. a fixed computer and a fixed programming language is chosen to evaluate all these algorithms. Python language and a laptop with a Core i3 (third generation Intel) CPU and 4GB of RAM and a 240GB SSD hard drive have been selected to do so. Therefore, all performance tests have been done on Python and the selected laptop. The important point is that the laptop does not have a very powerful processor and a very high capacity RAM, so the principle of choosing an algorithm will naturally be observed because the laptop will not be able to do it if there are heavy calculations.

Table 4.4. Test results of the top four machine learning algorithms

Algorithm	**K-Nearest Neighbors**		**Support Vector Machine**		**Decision Tree**		**Logistic Regression**		**Random Forest**	
Resident	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**
Accuracy	0.68	0.78	0.73	0.78	**0.86**	**0.95**	0.33	0.52	0.87	0.95
Precision	0.61	0.67	0.65	0.67	**0.79**	**0.93**	0.01	0.02	0.76	0.92
Recall	0.47	0.50	0.48	0.50	**0.75**	**0.90**	0.04	0.04	0.75	0.90
F1-score	0.51	0.56	0.53	0.56	**0.76**	**0.91**	0.02	0.03	0.75	0.91
Algorithm	**LightGBM**		**Gradient Boosting**		**CatBoost**		**XGBoost**		**Gaussian NB**	
Resident	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**
Accuracy	0.60	0.32	0.86	0.91	0.61	0.77	0.83	0.89	0.51	0.64
Precision	0.41	0.10	0.85	0.88	0.30	0.44	0.78	0.95	0.21	0.11
Recall	0.37	0.09	0.74	0.73	0.28	0.39	0.67	0.72	0.25	0.17
F1-score	0.37	0.09	0.77	0.77	0.27	0.40	0.71	0.79	0.20	0.12

4.10. Aras Dataset

d - The fourth step or the main step of testing the efficiency of the algorithm is on the dataset, so to enter this section, the dataset needs to be selected, i.e. using a dataset to test the performance tests on that dataset. In the next section, the dataset will be described, yet in this section, regardless of the nature of the dataset, performance results of machine learning algorithms will be touched upon, and a specialized explanation about the dataset will be given in the next section. As it can be seen in Table 4.5., in the selected dataset, there are two residents in the house named R1 and R2. The obtained results are the average results of the residents of the houses, and all the results of the evaluations of machine learning algorithms have been summarized in Table 4.4. Furthermore, performance test results of the selected algorithms which were able to predict the behavior of the dwellers of the house have very accurately been stated. In the following, Aras database will be explained with the help of the explanations in the work related to Aras database by Hande Alerndar et al.

Therefore, in that study, they offer the details of ARAS (Activity Recognition with Ambient Sensing), human activity recognition datasets collected from two real homes with multiple residents during 2 months.

Aras datasets have ground truth labels for 27 diverse activities. Each house was equipped with 20 binary sensors of various types which communicate wirelessly using ZigBee

protocol. A complete month of information containing the sensor data and the activity labels for both residents was collected from each home resulting in a total of 2 months' data.

As mentioned in the previous section, one of the most important steps in implementing and testing the proposed method is to select the appropriate dataset. A dataset is needed to implement the proposed method.

It is necessary to state again that no changes have been made in Aras dataset or datasets, it has been taken into account that all the available data has been extracted correctly, and smart environments and humans have not played any roles in changing it.

Several categories of available datasets have been touched upon in this field so far. Through examining the studies, one can find that the structure and type of dataset has a direct impact on all stages of the proposed methods, and even some very accurate methods may not be applicable at all due to the lack of a proper dataset. In choosing the dataset, it is very important to pay attention to the privacy of the residents of the house.

One of the reasons for selecting Aras dataset is to observe the privacy of the residents of the house very carefully. Due to the use of sensors, Aras dataset has observed a very high level of privacy. This is not the only reason for choosing Aras dataset. In addition to respecting the privacy of residents and not using cameras, Aras dataset is a real dataset which means that unlike some datasets, it is not a simulated or unreal dataset, yet is collected from people's real life.

In this chapter, all the implementation steps were reviewed, and the concepts of sub-NNCs, daily-NNCs and main-NNCs were delineated. Moreover, ways of calculating each of them were taught and methods of finding anomalies were explained.

It was stated that anomalies are not judged by unusual behavior, and the goal is to scan residents' behavior; in addition, it is emphasized that a set of behavior was looked at in this method. Finally, in this chapter, after introducing the algorithms available for use inside sub-NNCs and testing the performance of machine learning algorithms, Aras database was described and the reasons for choosing this database were explained.

5. EFFORTS TO INCREASE PRODUCTIVITY AND ACCURACY

5.1. Examining Possible Methods to Improve

In the previous chapter, the approach was described quite clearly and was explained how anomalies can be found by abnormally increasing or decreasing changes in the behavior of the residents of the house. In this way, it was decided that the decision criterion is based on intelligent comparison of main-NNC with daily-NNC, and it was understood that both daily-NNC and main-NNC are calculated by sub-NNCs. Hence, the most important part is sub-NNCs. Sub-NNCs are divided into two groups based on the type of rules inside them. The first group are the rules calculated based on statistical techniques, and the second group also have rules of computing which are done inside the machine learning technique.

In evaluating the results, the accuracy of the method is up to 75% in all evaluation parameters, modes, and conditions. This number is a relatively good accuracy for an anomaly warning system, yet it is emphasized that in the worst case, the accuracy of the method is over 75%. In some cases, the accuracy of this approach goes up to 96%. In this chapter, it is examined whether there is a need to increase the accuracy of the approach, and to increase the accuracy of the approach, what methods can be used and what solutions are available.

After ways of improving performance of the approach were reviewed, several different methods were come up with, which will be described in this chapter. Yet the remarkable point is that the first step to increasing the accuracy of performance should be to improve the accuracy of sub-NNCs' results, since, as in the previous chapter, a direct relationship between increasing the accuracy of sub-NNCs' performance and the accuracy of the entire approach was seen. Therefore, in order to increase the accuracy, the accuracy of the output results of sub-NNCs is attempted to increase first.

a - The first method for improvement is related to the dataset. Whether or not changes in the dataset could increase the accuracy of anomaly detection were examined. Clearing the data is a common way to improve performance. Changes in the dataset can be examined in two ways. The first type is in terms of data in the dataset, i.e. it is checked whether there is any wrong data or noise in the dataset. The second type is to modify the structure

and format the dataset; in other words, the same dataset information with a new structure or format should be provided to sub-NNCs. In contrast to the two ways of the above method, Aras dataset does not need any kind of modification and the information and the type of dataset structure is trusted. However, this trust has not been established without evaluation. In Aras dataset, only one column is considered for the moment of behavior. This means that seconds are stored in this column, although the time of behavior occurrence is also considered differently in Aras dataset and there is no need to do so. Thus, it was finally concluded that the desired result would not be achieved by changing Aras dataset, so, unfortunately, the first method to improve performance and upgrade was not useful.

b - The second method is related to the dataset which comes down to increasing the number of sensors in the smart home. At the very beginning of this method, it is not attempted to make changes in home equipment and sensors in the smart home; moreover, the goal is to increase the accuracy of the method without imposing a cost, and since the number of Aras dataset's data columns is satisfactory, the method will be ignored. Of course, it is worth mentioning that in houses where there is a possibility of increase or change in the sensors, it must be checked because a change in the position of the sensor can have a very crucial impact on the results and performance accuracy. Although this method did not work in this thesis, it will undoubtedly work in a real smart home, and the accuracy of detecting abnormal behavior can be increased by changing the number, type and location of sensors.

c - Another method which may significantly affect the accuracy of anomaly detection is expressed, and the interesting point here is that the third method is also related to the dataset yet in an indirect way. In this method, the use of thermal cameras or ordinary fixed cameras inside the house is proposed, and the detection accuracy is attempted to be enhanced by using the data obtained from the cameras. The use of cameras and image processing techniques is very effective, and can greatly improve the accuracy of correct anomaly detection, yet this method has two major drawbacks. The first downside is that in the previous chapter it was decided to use Aras dataset which means that smart homes do not have cameras and hence camera information cannot be accessed. On the other hand, changes cannot be made in the houses, and the most important disadvantage of this method is putting the comfort and convenience of home residents at risk; in fact, most

people are reluctant to insert a camera in their place of residence. Therefore, the theory of using information obtained from image processing is almost impossible in this master's thesis, at least not those smart homes in which the residents' privacy is important (Ramapatruni et al. 2019).

d - Nowadays, with the advent of artificial intelligence and related sciences, there is a great variety of intelligent assistants who are trying to make life easier for humans. In some of these smart assistants, the user who uses it is asked for help to increase accuracy and intelligence. For example, some smart time prediction software at the beginning of the road asks the user what means of transportation he or she will use on the route. The idea can be utilized to increase the accuracy of diagnosing anomalies through getting help from the residents of the house. This is a common method to increase the accuracy of the residents to ask for help in some cases. Of course, this method can lead to a lot of questions on the first days, but as the intelligence of the approach enhances, the number of questions the resident asks becomes less and less. This method can be very effective, but, unfortunately, at the moment one cannot count on it. To start with, the database is fixed, no information can be added to it, and it is not updated every day. Secondly, there is no connection with the residents of the house whatsoever. As a result, this method can be useful in real homes, yet it is ignored according to the circumstances.

All the above ways have been studied scientifically and practically, and it was concluded that they are impossible due to the conditions and data available in this master thesis. Of course, in some cases they would be possible in real smart homes. However, it should be noted that those cases, though feasible to implement, is in conflict with the policy of this approach. This means in terms of using cameras and processed files and analyzed images recorded from cameras, it would not be ethical because in previous chapters it was stated that using cameras was out of the question, and the most important principle is the house residents' mental comfort and privacy.

Regarding the last method, because of having no other connection with the residents, a more detailed study could not be done in this master thesis, but in other studies, one of the very good ways to improve the intelligence of a software is to establish intelligent communication with the user. Now the important question is whether there is another way

to improve the accuracy of the method. In the next section, the accuracy of the method is aimed to be improved from two angles (Kanev et al. 2017; Al Baalbaki et al. 2015).

After retrying and re-examining, new solutions were designed and are going to be explored. Two methods will be used to increase the accuracy of anomaly detection in the method. The first way will be to make an attempt to change the way daily-NCCs and main-NNCs are formed and thus increase accuracy with this change. Different methods can be tested to form daily-NNC and main-NNC. For example, regarding the calculation of the main-NNC, instead of calculating it from the simple average of daily-NNCs of the past days, one can gradually reduce the role of the days of the previous months in determining the main-NNC, i.e. the value of daily-NNC of the previous day will be different from daily-NNC of ninety days ago. These coefficients can be such that a larger impact factor for daily-NNCs of recent days and a smaller impact factor of the previous days are taken into consideration. With this trend, the main-NNC will always be up-to-date and will act more accurately.

Another example of this group to increase the accuracy of anomaly detection can be that sub-NNCs get a coefficient on the same day before the formation of daily-NNCs. It is certain that the output of all sub-NNCs is not equally important and effective, so a coefficient is defined which is called the coefficient of importance of the output of sub-NNC. Then, more important sub-NNCs will play a more significant role in the formation of daily-NNCs, and will naturally have a more major role in the main-NNC.

Examples of the first way to increase the accuracy of anomaly detection have been looked at, but unfortunately, in the final result, a maximum improvement of 5% was obtained at best which is experimental and cannot be definitively expressed. Therefore, it is very important to mention this point that it was not possible to be able to study the methods mentioned in the first way in detail in this research. The reason that no result can be obtained is, firstly, as more days go by, the coefficients may play a role in the success rate and one needs to look at how effective they are after a few months.

Secondly, in order to determine the importance of sub-NNCs, in the early days, the resident of the house needs to evaluate the accuracy of the diagnosis of his or her behavior by sub-NNCs. The resident himself or herself must determine which sub-NNC is more significant and effective. Hence, in the real smart home and in the early months, it is

necessary to establish an intelligent relationship with the residents of the house in the smart home with anomaly detection assistant. Moreover, an assistant with the guidance of the house residents to determine which sub-NNC has acted more accurately regarding their behavior is needed. This means that the importance and accuracy of sub-NNCs should be determined by the residents themselves.

As a result, it was concluded that it is true that changes towards calculations and comparisons can help the accuracy of the diagnosis of anomalies, but due to certain limitations in this master thesis, the methods mentioned in the previous paragraphs cannot help much to increase the accuracy of anomaly recognition. Thus, one needs to find a way to improve accuracy which can be implemented and used with the existing data and facilities. In addition, it provides a tangible result meaning that it could clearly be stated what percentage of the approach has improved, or what role it has played in improving the method.

The first way could initially be helpful in acting a little more accurately and intelligently, but because of the aforementioned problems, it was not convincing to end the efforts to improve at that point. A way other than the previous one is needed; a new way which would not lead to the previous problems. In the new solution, to increase the accuracy of anomaly detection, one should only focus on sub-NNCs, and to enhance the accuracy of anomaly detection, the diversity of sub-NNCs is reviewed. In other words, as it was witnessed in the previous chapter, there were only two sub-NNC groups. Now, the number and type of these two groups is attempted to be altered. To put it briefly and clearly, with the help of two different methods an effort is made to improve the approach in diagnosing abnormalities in the residents' behavior. In the following, two methods will be examined to increase the diversity and number of sub-NNCs.

5.2. New Models' Definition

a – As it was seen in the previous chapter, to diagnose abnormalities in the behavior of the residents of the house, sub-NNCs were defined. In this section, the accuracy of anomaly diagnosis is aimed to increase. After examining different methods, it was concluded that there exists no better solution than creating new sub-NNCs to improve the accuracy of anomaly diagnosis. There were two groups of sub-NNCs; the first group

moving forward based on statistical calculations, and the second were sub-NNCs based on machine learning algorithms. It was said in previous chapters that the best machine learning algorithm is selected and utilized within the second group of sub-NNCs.

This method is proposed to divide the second group of sub-NNCs into two or even three subgroups to increase the accuracy of diagnosing the anomaly. This means that machine learning-based sub-NNCs will be available in two subgroups; one subgroup will be the same as former sub-NNCs, and a new subgroup will be implemented with another machine learning algorithm. Therefore, all the rules will be executed at once with the help of the algorithm with the best results. Due to various reasons 'decision tree algorithm' was selected, and an algorithm from Table 5.1. was chosen once again and thus new sub-NNCs were implemented by that algorithm. Indeed, it was stated that there can exist two auxiliary algorithms and three subgroups in the second group. Yet algorithm will be selected this time and, if necessary, another algorithm will be chosen. As it was described in the algorithms in Table 5.1., 'decision tree algorithm' has already been selected for its advantages such as good accuracy, speed of operation, and very accurate results in predicting the occupants' behavior of the house. Now, using an auxiliary algorithm is attempted in addition to 'decision tree algorithm.' Selecting this algorithm can be based only on the accuracy of predicting the residents' behavior, or else on the estimated speed of response. It should be noted that which algorithm works faster on average would be significant. Although this criterion cannot be very accurate, one is able to estimate which of those algorithms works faster through several tests and evaluations (Tapia et al. 2004; Ramapatruni et al. 2019).

Table 5.1. Test results of the top four machine learning algorithms

Algorithm	**Gradient Boosting**		**Random Forest**		**Decision Tree**		**XGBoost**	
Resident	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**	**R1**	**R2**
Accuracy	**0.86**	**0.91**	**0.87**	**0.95**	**0.86**	**0.95**	**0.83**	**0.89**
Precision	**0.85**	**0.88**	**0.76**	**0.92**	**0.79**	**0.93**	**0.78**	**0.95**
Recall	**0.74**	**0.73**	**0.75**	**0.90**	**0.75**	**0.90**	**0.67**	**0.72**
F1-score	**0.77**	**0.77**	**0.75**	**0.91**	**0.76**	**0.91**	**0.71**	**0.79**

The main advantage of this method is its simple implementation to improve the diagnosis of abnormalities. In this method, no special and complex new task is carried out, and only

‘decision tree algorithm’ in the previous sub-NNCs is replaced with ‘Random Forest algorithm’ or ‘Gradient Boosting algorithm.’ On the downside, one can refer to the imposition of additional calculations on smart home computing center, and ask whether its computing and processing center is capable of performing new calculations, or by increasing the volume of calculations, they may not be able to do it at all.

A laptop for smart home computing center has been considered for this study. Acceptable hardware was selected considering that most smart homes may not have a very professional computing system. In addition to the previous sub-NNCs, they are also tested based on ‘Random Forest algorithm.’ As a result, the first major change is that computing time increases dramatically to the point where, on average, up to three times more computational time is spent. Of course, it should be noted that during the execution of the program, the laptop RAM is completely involved in the calculations and thus can find no time for other calculations in the smart home even for a few short moments.

After ‘Random Forest algorithm,’ ‘Gradient Boosting algorithm’ is checked. The latter’s results are very similar to the former’s, with the difference of a little faster results yet with a little less accuracy. Although the accuracy of prediction in ‘Gradient Boosting algorithm’ is a little less than ‘Random Forest algorithm,’ they do not differ much. However, the main problem both algorithms have is that they almost completely engage smart home computing center in their calculations, i.e. when those algorithms are running, no computational and processing can be practically done.

Using this method can increase the accuracy by about 8% on Aras dataset at best, but the maximum accuracy does not increase, which means that the resident’s behavior of the house can never be predicted with an accuracy above 96%. Defining sub-NNCs with new machine learning algorithms is very simple and efficient, so the first way to obtain a positive result is to define a subgroup for machine learning-based sub-NNCs. Of course, there exists a laptop with a capacity of 4 GB which was considered to implement this approach, but it is recommended that at least 8 GB RAM is utilized to perform calculations at a reasonable time, and if this laptop exists in smart home computing center, it is not recommended to use the two machine learning algorithms at all. Therefore, before making any decisions, the computing hardware of the smart home should be examined.

5.3. Deep Learning

b - The last method is very similar to the previous method. In this method, a new group of sub-NNCs is defined. This new group does not use statistical methods or machine learning algorithms. Inside these sub-NNCs are rules similar to the rules of the second group of sub-NNCs, except that they use deep learning algorithms instead of machine learning algorithms to predict the behavior of the residents. As it is known, in recent years, deep learning algorithms have made great strides, and have provided intelligent decision-making power to intelligent systems. Therefore, in this section, a new group is attempted to be created with the help of 'Deep Learning(DL)' which is capable of more accurately predicting the residents' behavior of the house. As it was mentioned, the third group of sub-NNCs in all rules is similar to machine learning-based sub-NNCs. The only discrepancy is that instead of using 'decision tree algorithm,' simple 'Neural Network(NN)' model is utilized in the behavior of the dwellers of the house (Mehr et al. 2016).

As it is known, ARAS dataset provides information about two people's activities inhabiting a house. This information is gathered by placing sensors of different types in the house and recoding the measurements in one month period.

Considering the fact that most routine activities of a person in the house and at any moment in time could be heavily influenced by their past activities, a simple 'Neural Network' model to detect abnormal behavior of the person is created and trained in this new method. A very basic example is that one can predict a person should be sleeping based on the fact that he or she has not slept for a long time, or should be eating if they have not eaten recently. If that is not the case, then it could indicate a health-threatening behavior, and the more this behavior continues, the more danger it poses to that person (VandeWeerd et al. 2020; Oguntala et al. 2019).

This model could be very useful, for example, to create an automated system to monitor elderly people's behavior living alone, and if the system detects a life-threating problem, then proper actions could be done. Hence, the goal is to create a simple model and learn the relationship between current daily activities of a person, for instance sleeping, with other activities of the same person, or even other inhabitants of the house in the last 24 hours (Galvão et al. 2017).

To capture the information of the sequence of activities in the last 24 hours, in a simple model— instead of using 'Recurrent Neural Network' which has its complexities, especially in the training process— a simple multilayer perceptron network was used, but the sequence of activities was given as input. Thus, every input vector at each moment includes activities of the last 24 hours.

An activity vector defines the average activity in the house for the duration of 60 seconds. It includes the values of the sensors placed within the furniture of the house, like when the couch or fridge is used as well as the explicit-activity-type of individual inhabitants of the house which is encoded as one-hot vector. With every sample of activity vector, each person's activity is explicitly understood (two one-hot encoded explicit-activity-type vector) in addition to the devices used in the house. In this dataset, every house has exactly 20 sensors and two inhabitants, and for each one, there are 27 different explicit activity types.

Therefore, encoding the activity type of each person as one-hot vector means that the explicit-activity-type vector has a size 27. An activity vector has a size 74 as it includes exactly 20 values for the sensors plus two explicit-activity-type each in a size 27 (Alemdar et al. 2013).

The input vector includes the sequence of activity vectors in the last 24 hours which is exactly 1440 activity vectors concatenated together without the explicit-activity-type for the last step which is used as the label to predict.

The output of the network is the explicit activity type predicted for every person encoded as one-hot vector. Thus, the output vector has also a size 2*27 which is exactly two explicit-activity-type vectors for every inhabitant of the house.

The model consists of 5 densely connected hidden layers including 50, 100, 500, 100 and 50 neurons respectively. Dropout layers were placed randomly to reduce the chance of overfitting. Hidden layers use 'Relu' activation and the output layer has 'Softmax' activation as Multi-class classification is being attempted. Instead of the classical stochastic gradient descent procedure, 'Adam' optimizer was used along with common Mean Squared Error loss function and a learning rate of 0.001 (Jethanandani et al. 2019).

The accuracy was managed to decrease drastically in the last attempt, but again the problem of needing a more powerful smart home management center is faced. At least 16 GB of RAM is needed to perform the calculations of the new method.

Certainly, as the price of hardware has decreased and become more available in the last several years due to the development of hardware component manufacturing industry, smart homes will soon be seen in the form of very advanced computing hardware.

The above methods showed that a more advanced computing core is needed to increase accuracy. In fact, more than a new database or new methods of machine learning or deep learning, smart home management center with powerful and advanced hardware is required. On the other hand, a very important point is that with the trend machine learning and deep learning algorithms are evolving and improving every day, it will soon be possible to create algorithms responding to fewer calculations and thus more efficient algorithms are hoped to be seen as well.

6. CONCLUSIONS

6.1. Assumptions and Limitations

Facing limitations in any scientific and real work is an undeniable thing. These limitations can have different types depending on the type of research. Ignoring the limitations of the real environment is perhaps the greatest disadvantage of research and scientific work; therefore, it is very important and necessary to try to identify the limitations in each study from the beginning. In the first stage, the limitations should be attempted to be removed in the real situation. If they are not removed and covered, every effort should be made to manage them, so that they do not hinder the practical implementation of the research.

In this master's thesis, all the limitations were attempted to be taken into account at all stages, and efforts were made to study in a completely real environment instead of implementing it in a laboratory and unreal environment. In the pre-study phase of this thesis, what constraints currently exist in the real world for the topic at hand was discussed. Two general categories were come across; the first of which was the most common limitations of hardware sector, most of which were very expensive to cover, and the second category was related to software and algorithms. Of course, with the development of smart homes, algorithms, and software related to smart homes, these restrictions are becoming less and less, yet as it was mentioned, hardware-related restrictions are much slower and more expensive.

In this study, an ideal smart home had been considered with a set of ideal hardware and software to start expressing the problem. Yet a real dataset and a set of common hardware in the home were forced to be touched upon. Thus, the limitations consisted of two categories: the first category was related to Aras database and the information received about the conditions of the house and its occupants, and in the second category, certain limitations were set, so the study continued according to those assumptions. For example, a very common laptop with a 4 GB of RAM was utilized in all the steps of running the code, while a computer server with a 128 GB of RAM could be used. The reason for this was that one had to look at what percentage of today's popular smart home computing center has such an advanced computing center. Since most smart homes today do not have very powerful processors, a very simple laptop was also used to implement work steps.

Certainly, many of the limitations of this thesis will most likely be displaced in a few years. Different types of sensors are produced every day in the world which are capable of producing much more accurate information about the condition of the house and its occupants and are delivered to researchers. On the other hand, scientists working in related fields provide more accurate and faster prediction algorithms every day. Moreover, the combination of these two trends will soon provide homes and environments incomparable in all respects with today's smart environments. Therefore, it can be said that this thesis' limitations will be lifted soon.

6.2. Future Studies

Accurately predicting human behavior during the day is certainly one of the most complex tasks, but it should be noted that firstly, accurate prediction is not attempted, and secondly, the goal is just to design a method which predicts human behavior with acceptable accuracy and finds any abnormalities, so in the future, increasing the accuracy and efficiency of this process will only be sought. As stated at the beginning of this thesis, it is endeavored to take steps with the available facilities and capabilities of this issue, and therefore the first thing which can be decided to do in the continuation of this study is that in both parts the steps were limited to go further in the discussion of datasets, to examine more existing datasets, and to try to take advantage of new datasets providing researchers with more information. In the next step, the brain of smart environments can be given more power, i.e. to consider a more advanced computing system in the computing center (a computer having more and faster computing and processing power). In the design and selection phase of the algorithm, one should not think about whether the computing center has the capability to calculate and manage this algorithm or not, but rather should focus more on the accuracy of the proposed algorithm.

After taking those two steps, having a new perspective on the existing algorithms and techniques would be desirable, and most likely, combining machine learning techniques with new neural networks would seem wonderful. The neural network can help researchers a lot in increasing the accuracy of predicting and recognizing people's behavior. Of course, it cannot be said that this attitude will definitely result in better outcomes, yet it will most likely provide valuable experiences for those who are interested in the topic.

Before drawing a conclusion, if finding support from an industry or a university to collect a real dataset takes place, designing a real smart home and using the data of that house would be fantastic. The main goal is to test and evaluate the proposed algorithm immediately, and receive the residents' opinions to check the accuracy of the algorithm. Of course, everyone would be very motivated to join if one day a sample or pilot of such studies performed somewhere.

6.3. Summary

In the last chapter of this master's thesis a general summary will be given, but this does not mean that the purpose of this section is to provide a summary of previous chapters. In this section, a conclusion is sought to be drawn as well as providing a brief summary of all the taken steps. It is hoped that these studies will be useful for those who want to continue their research in this field. Summarizing all the efforts is sought to present the approach, yet for a more accurate understanding, it is necessary to study all the chapters. In the following, two suggestions will be made; one concerning changing the anomaly check period, and the other about the data set. It should be noted that in this study, residents' behavior of the house was examined every 24 hours, and the dataset used was 'Aras dataset.'

In the first part, it is necessary to state the reason behind choosing this topic in order to precisely know what goal is aimed to achieve. As it was stated in the previous sections, the reasons for the need to conduct research in this area are very diverse, and with the significant growth in the number of smart homes around the world, attention to this issue has become greater. The purpose in presenting this topic is being able to express an approach to identify normal and abnormal behavior of residents. An approach is tried to be created which can help one find people's abnormal behavior and create an assistant for smart home management center, so that if there is an abnormality in the behavior of the dweller, it can be detected and treated a little more intelligently.

Despite the reasons for the rapid change of traditional homes to smart homes, and the logic behind the increase in the number of single-resident homes are very important, attention is not paid to them making it even more urgent to turn smart homes into a more pleasant, easier, and safer environment. It is noteworthy at all stages that this approach must be fully practical and applicable in existing smart homes. Moreover, as it was stated,

with the growth of single-resident homes, it becomes even more important to recognize if the behavior of the resident changes. In addition, recognizing abnormalities in the behavior of the resident of the house is necessary to create a safe environment for the resident. For example, consider an elderly person living alone in his or her home. If his or her behavior becomes abnormal one day, it is very crucial that his family or friends find out about his or her abnormal behavior such as sleeping out of the ordinary, going to the bathroom all too often, or not even going to the kitchen to eat.

Therefore, it is crucial to try to come up with an approach to enhance the smart communication of smart home management center because it is capable of making more accurate and intelligent decisions via having an accurate understanding of home conditions and its residents. Moreover, for the smart home management center, predicting behavior and detecting anomalies in the behavior of home residents is the first and most important step in finding an understanding of home.

Before giving an overview of the approach, it should be noted that although one of the incentives for choosing this topic is single-family smart homes, the offered approach can be implemented in all smart homes with any number of residents, so it does not mean that the proposed approach is limited to the number of members of the house, and can only be implemented in single-resident homes. It is also noteworthy that in non-single-resident homes, other residents' behavior can be taken into account in predicting the resident's behavior, and very similar patterns may be found. For example, in a house where a couple lives together, the two residents usually eat at the same time, and when one dweller is asleep, the other is asleep as well.

To begin with, considering that the purpose of this master's thesis is to provide an approach to detect abnormalities in the residents' behavior of a smart environment, articles and studies in this field have been studied, and results and experiences of previous studies and research have been sought to be used to reach a design of the solution, and learn all the advantages and disadvantages of existing methods. After this step, a series of principles have been set not to attempt to violate those principles. For example, it was said that the amount of calculations need to be done to achieve the desired result, should not be too heavy and can be implemented in smart home computing processing center in

the world. In some studies, some methods have been proposed which require very heavy and complex calculations.

It was also stated that protecting the privacy of residents in smart homes is a fundamental principle for researchers, and this principle will be never violated, even if it leads to an increase in the accuracy of the approach; moreover, in this approach, putting the resident's privacy at risk was avoided. For example, the presence of a camera in the house will most likely not make the residents of the house feel comfortable, so it was decided not to even use the camera data, which means that in this approach, there should not be an urgent need for a camera in the home. After summarizing previous studies in the field of human behavior detection and abnormalities in human behavior, the approach was started to be designed. In this way, all the steps were sought to be completed with repeated evaluations in the best possible way (Pan et al. 2016; Ali et al. 2017).

As it was explained in the previous sections, this study does not aim at designing software or providing a computational method, but rather it aims to provide an approach which allows the researchers to use existing methods, and discover abnormalities in the behavior of residents. To say that this is an approach does not mean that it is purely theoretical, and the possibility of practical implementation was having not been proposed. In this thesis, practical steps of this approach were also reviewed and evaluated as much as possible. In this way, the concept of NNC was expressed and stated that through using computational methods, one could try to turn the existing irregularities in the behavior of each resident of the house into an order, i.e. to create a pattern of each person's behavior without paying attention to anything. If a change is seen in a person's pattern of behavior, it can be concluded that the behavior of the house resident has become abnormal. One of the most important features of this kind of outlook to find anomaly is that one can find anomaly even for people who perform dozens of unusual behavior every day. It is very interesting that in this approach, even if an irregular person becomes a regular being one day, that day will be an unusual day for that person, and his or her behavior are considered to be abnormal, so a pattern of behavior can be created for each resident, and by comparing that pattern with the new day, the occurrence or non-occurrence of anomalies in the behavior of the resident of the house can be found.

Another advantage of this method is that all the behavior of the resident of the house is looked upon, and the judgment is based on a set of behavior. Of course, if smart home management center needs it, a small change can be made in the implementation making the approach sensitive even for minor anomalies. In other words, in the implementation part, this approach is changed, so that it does not wait for all the behavior to be evaluated and the final result be stated. An alarm is created inside each sub-NNC that, upon seeing any abnormality in the resident's behavior, issues a report for investigation by the residents themselves.

In the previous chapters, after describing how to find the anomaly, different tests were carried out, and after achieving acceptable results in the accuracy and speed of anomaly diagnosis, the step by step approach was presented. However, in the next chapter, after presenting the approach, Actions were taken to increase the accuracy of anomaly detection and provide a more accurate approach by improving the accuracy of predicting the behavior of the dweller of the house.

It is very interesting that when it was sought to increase the accuracy of anomaly detection and prediction of resident behavior, an important point was reached which was how flexible this approach could be. Moreover, how easily can one greatly increase the accuracy of recognizing the behavior of the residents of the house by using new methods such as deep learning is worthwhile, although due to the sharp increase in the computational load in smart home management center, the same amount of accuracy in predicting resident behavior would seem satisfactory (Yachir et al. 2019; Natani et al. 2019; Mehr et al. 2016).

In the improving efforts, it was seen quite practically how dynamic this approach is in diagnosing abnormalities. Moreover, perhaps the best part is that over time, as new methods emerge, they can be applied internally and used to provide a more accurate system. As it is evident from the rapid development of computing technology, there will soon be smart homes which have been improved in terms of hardware, and will have a very powerful computing center, and a new generation of processors will calculate all the required calculations in the shortest time. Therefore, it is hoped that all sub-NNCs be found simultaneously in smart homes in the near future.

It was stated in the approach that attention will be paid to all the details in the behavior of the residents and the condition of sensors, but the principle is the general behavior of the residents of the house during the day. In fact, a quick decision will not be made with regarding the behavior, although certain behavior is very important to the researchers in diagnosing abnormalities. Of course, in this section it is aimed to state that if 24 hours is a long time to diagnose anomalies, a day can be divided into three or four parts or even more, or else even all these calculations can be performed at the end of each hour and anomalies be checked (Yoshimizu et al. 2014).

Changing time interval from 24 hours to less, allows one to pay more attention to detail, but due to the sharp increase in the amount of the required calculations, smart home computing center will be seriously involved in its calculations and will not practically do any other processing or calculations and hence it is not possible to implement and perform those calculations in the center. It should be noted that the reason for this restriction is that as far as it was examined, current smart homes do not have a very professional computing center and therefore the behavior of the residents could only be examined once every 24 hours.

Finally, in the case of database, it is very important to state that this study can be much more powerful in the database section; it means that changes could be utilized in the dataset to increase productivity and detection accuracy, yet the aims were not to focus on the data in this approach, and only the existing datasets were used even without the slightest change, but to implement the approach in the future, real datasets of accurate sensors from real smart homes are desperately needed. Fortunately, companies have reached a very good level in the production of sensors and modern equipment nowadays. Moreover, it is hoped that better datasets in this field will soon be collected from real homes and not in an experimental environment which will allow the researchers to make more accurate decisions. The fact is that there were a lot of limitations in terms of dataset in this master's thesis, yet given the conditions of today's smart homes, Aras is a completely acceptable dataset among all other datasets.

Although the goal was just to provide an approach, it was sought to implement all the steps accurately. Each step of this approach was examined using Python programming language, and it was tried to make sure that the approach is not just a theoretical and

contemplative approach, and that it can be implemented and used in smart homes. The results which are generally stated have always been sought to be accurate and almost all evaluations have been done several times so that there is no doubt about the result, but it must be said that there is a possibility of error in all evaluations and tests and I apologize for all the shortcomings which may have happened by accident. Finally, I would like to thank the professors and friends who have worked in this field before us. It is also hoped that the efforts can be of great help in this area of research. Special thanks to the dear professors and friends who help and accompany us in this path.

REFERENCES

Al Baalbaki, B., Pacheco, J., Tunc, C., Hariri, S., Al-Nashif, Y. (2015). Anomaly behavior analysis system for ZigBee in smart buildings. *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 1-4.

Alemdar, H. (2014). Multi-resident human behaviour identification in ambient assisted living environments. *Proceedings of the 16th International Conference on Multimodal Interaction*, 369-73.

Alemdar, H. (2015). Human activity recognition with wireless sensor networks using machine learning. PhD Thesis, Bogaziçi University, Istanbul, 387373.

Alemdar, H., Ertan, H., Incel, O. D., Ersoy, C. (2013). ARAS human activity datasets in multiple homes with multiple residents. *7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 232-35.

Ali, W., Dustgeer, G., Awais, M., Shah, M. A. (2017). IoT based smart home: Security challenges, security requirements and solutions. *2017 23rd International Conference on Automation and Computing (ICAC)*, 1-6.

Ang, I. B. A., Salim, F. D., Hamilton, M. (2016). Human occupancy recognition with multivariate ambient sensors. *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, 1-6.

Aran, O., Sanchez-Cortes, D., Do, M.-T., Gatica-Perez, D. (2016). Anomaly detection in elderly daily behavior in ambient sensing environments. *International Workshop on Human Behavior Understanding*, Springer, 51-67.

Belapurkar, N., Shelke, S., Aksanli, B. (2018). The case for ambient sensing for human activity detection. *Proceedings of the 8th International Conference on the Internet of Things*, 1-3.

Bhargava, K., McManus, G., Ivanov, S. (2017). Fog-centric localization for ambient assisted living. *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1424-30.

Deep, S., Zheng, X., Karmakar, C., Yu, D., Hamey, L. G., Jin, J. (2019). A survey on anomalous behavior detection for elderly care using dense-sensing networks. *IEEE Communications Surveys & Tutorials*, 22: 352-70.

Dilraj, M., Nimmy, K., Sankaran, S. (2019). Towards behavioral profiling based anomaly detection for smart homes. *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, 1258-63.

Diraco, G., Leone, A., Caroppo, A., Siciliano, P. (2019). Deep Learning and Machine Learning Techniques for Change Detection in Behavior Monitoring. *AI* AAL@ AI* IA*, 38-50.

Ertan, H. (2013). Human activity recognition using wireless ambient sensor networks. Master Thesis, Bogaziçi University, Istanbul, 325528.

Fernandez-Carmona, M., Cosar, S., Coppola, C., Bellotto, N. (2017). Entropy-based abnormal activity detection fusing RGB-D and domotic sensors. *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 42-48.

Galvão, Y. M., Albuquerque, V. A., Fernandes, B. J., Valença, M. J. (2017). Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1-6.

Ghahramani, A., Pantelic, J., Lindberg, C., Mehl, M., Srinivasan, K., Gilligan, B., Arens, E. (2018). Learning occupants' workplace interactions from wearable and stationary ambient sensing systems, *Applied energy*, 230: 42-51.

Ghayvat, H., Awais, M., Pandya, S., Ren, H., Akbarzadeh, S., Chandra Mukhopadhyay, S., Chen, C., Gope, P., Chouhan, A., Chen, W. (2019). Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. *Sensors*, 19: 766.

Igwe, O. M., Wang, Y., Giakos, G. C. (2018). Activity learning and recognition using margin setting algorithm in smart homes. *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 653-58.

Jethanandani, M., Perumal, T., Chang, J.-R., Sharma, A., Bao, Y. (2019). Multi-Resident Activity Recognition using Multi-Label Classification in Ambient Sensing Smart Homes. *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 1-2.

Kanev, A., Nasteka, A., Bessonova, C., Nevmerzhitsky, D., Silaev, A., Efremov, A., Nikiforova, K. (2017). Anomaly detection in wireless sensor network of the "smart home" system. *2017 20th Conference of Open Innovations Association (FRUCT)*, 118-24.

Khan, M. A. A. H., Roy, N. (2017). Transact: Transfer learning enabled activity recognition. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 545-50.

Makonin, S., Bartram, L., Popowich, F. (2012). A smarter smart home: Case studies of ambient intelligence. *IEEE Pervasive Computing*, 12: 58-66.

Mehr, H. D., Polat, H., Cetin, A. (2016). Resident activity recognition in smart homes by using artificial neural networks. *2016 4th international istanbul smart grid congress and fair (ICSG)*, 1-5.

Natani, A., Sharma, A., Peruma, T., Sukhavasi, S. (2019). Deep learning for multi-resident activity recognition in ambient sensing smart homes. *2019 IEEE 8th Global conference on consumer electronics (GCCE)*, 340-41.

Nef, T., Urwyler, P., Büchler, M., Tarnanas, I., Stucki, R., Cazzoli, D., Müri, R., Mosimann, U. (2015). Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data. *Sensors*, 15: 11725-40.

Oguntala, G. A., Abd-Alhameed, R. A., Ali, N. T., Hu, Y.-F., Noras, J. M., Eya, N. N., Elfergani, I., Rodriguez, J. (2019). SmartWall: novel RFID-enabled ambient human activity recognition using machine learning for unobtrusive health monitoring. *IEEE Access*, 7: 68022-33.

Otebolaku, A., Enamamu, T., Alfoudi, A., Ikpehai, A., Marchang, J., Lee, G. M. (2020). Deep Sensing: Inertial and Ambient Sensing for Activity Context Recognition Using Deep Convolutional Neural Networks. *Sensors*, 20: 3803.

Ozen, M. O. (2013). Careact: An Adaptive And Context Aware Framework For Anomaly Detection In Ambient Assisted Living. Master Thesis, Bogaziçi University, Istanbul, 338819.

Pan, Z., Pacheco, J., Hariri, S. (2016). Anomaly behavior analysis for building automation systems. *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, 1-8.

Parvin, P., Chessa, S., Kaptein, M., Paternò, F. (2019). Personalized real-time anomaly detection and health feedback for older adults. *Journal of Ambient Intelligence and Smart Environments*, 11: 453-69.

Parvin, P., Chessa, S., Manca, M., Paterno, F. (2018). Real-time anomaly detection in elderly behavior with the support of task models. *Proceedings of the ACM on human-computer interaction*, 2: 1-18.

Ramapatruni, S., Narayanan, S. N., Mittal, S., Joshi, A., Joshi, K. (2019). Anomaly detection models for smart home security. *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, 19-24.

Shelke, S., Aksanli, B. (2019). Static and dynamic activity detection with ambient sensors in smart spaces. *Sensors*, 19: 804.

Stavropoulos, T. G., Meditskos, G., Andreadis, S., Avgerinakis, K., Adam, K., Kompatsiaris, I. (2020). Semantic event fusion of computer vision and ambient sensor data for activity recognition to support dementia care. *Journal of Ambient Intelligence and Humanized Computing*, 11: 3057-72.

Tapia, E. M., Intille, S. S., Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. *International conference on pervasive computing*, Springer, 158-75.

URL-1: https://www.dcube.co.uk/news/how-smart-home-technology-is-changing - architecture/, (Visit Date: 15 May 2020).

URL-2: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures, (Visit Date: 27 August 2020).

VandeWeerd, C., Yalcin, A., Aden-Buie, G., Wang, Y., Roberts, M., Mahser, N., Fnu, C., Fabiano, D. (2020). HomeSense: Design of an ambient home health and wellness monitoring platform for older adults. *Health and Technology*. 10: 1291-309.

Wang, W., Chen, J., Hong, T. (2018). Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings. *Automation in Construction*, 94: 233-43.

Yachir, A., Amamra, A., Djamaa, B., Zerrouki, A., and khierEddine Amour, A. (2019). A Deep learning and multimodal ambient sensing framework for human activity recognition. *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 101-05.

Yang, Z., Li, N., Becerik-Gerber, B., Orosz, M. (2014). A systematic approach to occupancy modeling in ambient sensor-rich buildings. *Simulation*, 90: 960-77.

Yoshimizu, K., Takemura, N., Iwai, Y., Sato, K. (2014). Multi-sensor-based ambient sensing system for the estimation of comfort/discomfort during desk work. *2014 IEEE/SICE International Symposium on System Integration*, 425-30.

Zhou, Y., Chen, X., Zomaya, A. Y., Wang, L., and Hu, S. (2015). A dynamic programming algorithm for leveraging probabilistic detection of energy theft in smart home. *IEEE Transactions on Emerging Topics in Computing*, 3: 502-13.

PERSONAL PUBLICATIONS AND ARTICLES

Seyedkazemi Ardebili, E., Eken, S., Küçük, K. (2020). Activity Recognition for Ambient Sensing Data and Rule Based Anomaly Detection. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 44: 379-82.

BIOGRAPHY

In 2011 he entered the state university, and in 2016, he graduated from the University of Mohaghegh Ardabili with a bachelor's degree in software engineering with a grade point average of 16.01 out of 20. In addition to studying, he started working in a field related to his field. After the university, he entered forced military service, during which was also in the office of the Plan and Budget Organization of the Islamic Republic of Iran. Due to his interest in the country and culture of Turkey, he decided to continue his studies in there. In 2018, he was accepted into Kocaeli University for a master's degree in computer engineering.