SMART HOME AND MACHINE LEARNING FOR MEDICAL SURVEILLANCE: CLASSIFICATION ALGORITHMS SURVEY

1LAMIAE ELOUTOUATE, 2ELOUAAI FATIHA, 3GIBET TANI HICHAM, 4BOUHORMA MOHAMMED

1, 2, 4Laboratory of Informatics Systems and Telecommunications (LIST), Departement of Computer Engineering, Faculty of Sciences and Technologies of Tangier, Abdelmalek Essaadi University, Morocco
3Laboratory of Advanced Sciences and Technologies (STA), Departement of Computer Science, Multidisciplinary Faculty of Larache, Abdelmalek Essaadi University, Morocco
E-mail: 1lamiae.eloutouate@uae.ac.ma, 2elouaaif@gmail.com, 3h.gibettani@uae.ac.ma, 4bouhorma@gmail.com

ABSTRACT

With the recent advancements on the computer-engineering field, the paradigm of smart home has been increasingly suggested as an empowering solution for various issues. Smart home employs the most novel technologies, such as wearable technologies, the Internet of Things (IoT), cloud computing and machine learning analysis capabilities to change the way we live. Accordingly, a smart home for medical surveillance would certainly reinforce the smart healthcare model, thus making healthcare system further accomplished, more comfortable and customizable. With the intent of creating a convenient smart home for medical surveillance, in this paper, we first introduce a novel architecture for a smart home aimed to monitor a patient specific health condition and update the health practitioner with the patient data. Then, we exhibit a comparative study of several machine-learning classification algorithms capable of classifying a patient arising health condition and ultimately decide whether to raise a concern notification, call for help or only log the information.

Keywords: Smart Home, Smart Healthcare, Machine Learning, Medical Surveillance, IoT

1. INTRODUCTION

There is no doubt that smart home technologies have known extensive developments thanks to the rise of wearable technologies, the Internet of Things (IoT), cloud computing and machine learning which have inspired many researchers and businesses to consider mobile and remote medical surveillance for people with health conditions, the elderly and people with reduced mobility [1]. In the same context, remote medical surveillance systems incorporate the activities of data gathering, processing and analysis. Health data can be collected from smart phones, wearables and specialized sensors such as temperature, blood pressure, electrocardiogram (ECG)… By concentrating the previous mentioned technologies and putting them to work together inside a patient home, we can create a healthy and sustainable environment that continuously monitors and assist a patient without the need of overcrowding hospitals and nurseries [2].

Numerous studies and researches are produced every year to feature the application of trending technologies such as IoT and cloud computing on healthcare. Among these studies, Dev Gupta et al. [3] proposed a smart device built using temperature, heartbeat and blood pressure sensors and combined with Arduino for monitoring the health parameters related to obese patients. The proposed device would continuously update the doctor with the patient medical status, which can help detect an emergency beforehand. The authors in [4] reviewed a diversity of relevant papers that tackles health monitoring and smart home technologies in order to determine the maturity of the proposed solution on leveraging the healthcare system. This study concluded that the status of technology fitness for smart home health monitoring is deficient. A. Rghioui et al. produced another study [5] on the development of a smart healthcare system, where they focused on providing a survey on IoT and its applicability in healthcare. Additionally, the authors presented some insights about the recent trending technologies and their possible benefits for smart healthcare. On paper [6], the authors introduced a
cloud based smart home environment for healthcare. By discussing their proposed architecture, the authors established a case study that demonstrate how contextual information, collected inside the smart home environment, can be integrated to health data to create medical exhaustive data that can assist caretakers on perceiving the health status of a specific patient. To the extent of our knowledge, all of the previous mentioned researches have discussed extensively the technological aspects of a smart home for health monitoring as well as how it can help advance the smart healthcare system. However, these studies did not manage to propose an optimal design for a smart home for medical surveillance and focused only on the element of collecting health data and presenting it to healthcare providers without further discussions.

This paper illustrates a proposed smart home design for medical surveillance that considers a centric mobile application connected to several wearable and sensor devices to monitor a patient health condition. The assembled data is then examined, analyzed and sent to the health practitioner in charge of the patient condition. The processing and analysis operation uses machine-learning algorithms to detect the severity of a condition, which can then decide to alert the patient, notify a family member, call for medical assistance or only log the information on the patient report. The paper also presents a comparison study of machine learning algorithms that aims to highlight the most efficient algorithm for data analysis and notification evaluation in case of a patient with medical heart condition or blood pressure condition.

The rest of this paper will be arranged as follows. Section 2 focus on presenting the proposed smart home for medical surveillance architecture and its different components. This section also includes the functioning process of the proposed smart home design. Section 3 includes the machine learning comparative study and its results for medical health monitoring and alerting. The paper is then concluded in section 4.

2. RESEARCH METHOD AND PROPOSED DESIGN

At first, we developed a generic and patient centric smart home architecture, shown in figure 1, which considers the following basic technologies:

**Wearables:** These are smart devices envisioned to improve people’s quality of life [7][8]. Respectively, the current improvement on this technology enables it to collect valuable data such as temperature, blood pressure, heartbeat. The proposed smart home architecture considers wearables as endpoint devices for data collection.

**Smart phone:** Smart phone and mainly mobile applications [9][10] have become a key component inside every home. This technology allows a user to gather information from various devices (such as TV, cameras, security locks...) along with advanced controlling and analysis capabilities. In our case, this essential item will be the central hub that gathers data from the other smart home devices, wearables and hand it over to the central analysis platform.

**Cloud Computing:** This computing model supply the users with the required information technology resources as services [11][12]. This third key element will host the major part of our proposed system for the smart home management, data processing and alerting agent. The primary objective of using cloud computing on the proposed smart home for medical surveillance is to reduce the energy consumption on the patient smart phone where only a lightweight application will be installed that access everything on the cloud central application.

![Figure 1. The Proposed Smart Home Architecture](image_url)

2.1 The proposed smart phone mobile application

In recent times, smart phones and mobile applications have emerged into a centric fact [13]. Along with the quick development of smart phone technologies, mobile applications started to cover practically everything in day-to-day activities: e-commerce, entertainment, education as well as healthcare. In view of smart phones and mobile applications characteristics from whom we can mention: mobility, connectivity, personalization and location awareness, today smart home architectures are built using mobile applications as executive devices.

The proposed mobile application on the envisioned smart home architecture acts as a patient
agenda, a notification management platform and a commuting endpoint. The application is lightweight where it accesses everything as web services (using JSON technology) on the main application placed on the cloud (figure 2):

- **Patient Agenda**: This the central user platform that let him check and follow a specific medical schedule planed and validated by the health practitioner.
- **Notification management platform**: The main application generates notifications on the main cloud application and send them to the patient smart phone. A notification is a sort of alert that point the user to a specific information on the main agenda (missing a check-up, medication…).
- **Commuting endpoint**: The patient mobile application aggregates data from wearables, sensors and transfers them to the main application on the cloud for processing.

### 2.2 The proposed cloud main application

Cloud computing is an ubiquitous computing model that enables users to improve information technology management and reduces operating risks [14][15]. Without citing the significance of cloud computing advantages on various fields, this innovative paradigm forms a solid podium for mobile applications.

From our perspective, using cloud computing is an essential aspect of the proposed smart home architecture that addresses the following concerns:

- Reducing the energy consumption on the user smart phone in order to extend the battery daily usability and life.
- Eliminating the need of investing on a new smart phone device with advanced performances, where the major and heavy processing and manipulations are executed on the cloud main application.

The proposed cloud main application is developed using JAVA (figure 3) and operates with MYSQL database to store and manage data. Additionally, the main application invokes Python scripts for data processing and analysis using machine-learning algorithms. Respectively, the proposed application incorporate the succeeding elements:

- The health practitioner administrative panel: it represents a command center for the health practitioner (doctor). This is where the patient medical condition is detailed and the particular controls are set to monitor the health condition of the patient. This panel also let the health practitioner filter the type of data to be gathered and stored for processing according to the patient condition.
- The patient agenda: This the administrative module for the patient agenda configuration. It gives the health practitioner the capability of personalizing the patient medical schedule according to his condition.
- Data processing and analysis: This is the most important component of the proposed smart home for medical surveillance. This crucial element utilizes machine-learning algorithms to help classify the patient health condition based on the collected data, thus it provides the whole system with a high level of intelligence that let it decide when to raise a concern about the monitored patient.
- The notification platform: This coordination segment organizes, distributes and archives notifications.

### 2.3 The proposed health practitioner endpoint

The health practitioner (doctor) is a principal actor on the proposed smart home medical surveillance design. Therefore, he is in charge of controlling, monitoring and evaluating the patient condition as well as constructing the patient medical agenda. To accomplish these tasks, the health practitioner (doctor) will use an endpoint application accessible via smart phone or a desktop computer device, whereas its fundamental features are:

- A dashboard to monitor the patient vital measurements and health overall condition centralized on the cloud main application.
- A primary interface to access the administrative panel to construct the patient medical condition along with all the settings regarding the condition severity and the required data labels for monitoring.
- An administrative tab to access the patient agenda module on the cloud for configuration.
- A notification tab that alerts the health practitioner in case of an emergency.

### 2.4 Data processing and analysis technics

Machine learning (ML) is a study field that encompass the procedures and approaches of analyzing and deducing patterns from large data sets and eventually creating a prediction model capable of making decisions of new data. Machine learning approaches or methods can be arranged on three categories [16][17]:

- Supervised learning: This machine-learning category represents a set of algorithms that
exploits an initial labeled data in order to make accurate predictions on any new given set of data. An example of the algorithms of this category are linear regression and Decision tree.

- Unsupervised learning: This second category is a collection of algorithms and techniques that help splitting data into subgroups like clustering. We can find K-means and K Nearest Neighbor as unsupervised leaning models.

- Reinforcement learning: This third category operates contrarily from supervised learning, where the algorithms of this method scores or rewards data based on the obtained output and reuses the actions that led to a successful output on the future predictions. Q-learning is an example of the algorithms used on this category.

By investigating the three machine learning major categories and the problematic in hand, we can clearly recognize that the proposed alerting system for the patient health condition is a classification issue, given that almost every medical data is already labeled. Nevertheless, choosing the correct algorithm or technic to implement for the proposed system is not an easy job as consequence to the many variables of the smart home design: Technologies, cost, performance... We can also add the precision as a major concern for a medical surveillance smart home. Consequently, a comparative study of the existing supervised learning algorithms is an essential constraint that would serve as a framework to select the best algorithm for an alerting system for the proposed medical surveillance smart home.

From the major machine learning classification algorithms [18][19][20], we can mention:
- Linear regression: this is a category of regression that uses the linear relationship between two variables by approximating it to a straight line [21].

- Logistic regression: this is a statistical method that characterise the relationships between a binary response variable and one or more explanatory variables, which can be categorical or numerical [22].

- Naïve Bayesian networks: Compactly represents and encodes joint distributions of random variables. This machine-learning algorithm utilizes conditional dependencies and independencies to eliminate unnecessary parameters [23].

- Multi-Layer perceptron (Neural Networks): this is a category of machine learning algorithms that mimics the human brain neural network. This type of neural network is organized in several layers within which, information flows from the input layer to the output layer (feedforward network) [16].

- Support vector machine: SVM belongs to the category of linear classifiers. SVM main target is to separate the data into classes using a border that is as "simple" as possible, so that the distance between the different groups of data and the border which separates them is maximum [23].

- K Nearest Neighbor: KNN algorithm uses an intelligent prediction method that consists on looking for the nearest neighbors (using Euclidean distance, or others) for each new input and choosing the class of majority neighbors [23].

- Decision tree: Decision tree algorithms are a category of trees used in data mining and business intelligence. They employ hierarchical representation of the data structure in the form of decisions sequences (tests) in order to make a prediction. Each input must be assigned to a class that is described by a set of variables, which are tested in the nodes of the tree. Testing takes place in internal nodes and decisions are made in leaf nodes [24].

3. RESULTS AND DISCUSSIONS

All the classification machine learning models cited on the previous section require a dataset for training. Accordingly and in the interest of better understanding and comparing these algorithms, we selected the cardiovascular diseases dataset produced by “Svetlana Ulianova” a Data Science Student at Ryerson University. The dataset is largely accessible via “Kaggle” and consists of 70,000 records of patients data. The dataset comprises the following parameters: Age (days), gender (1 for male and 2 for female), height (in centimeters), weight (in centimeters), arterial pressure high (integer), arterial pressure low (integer), cholesterol (1 for normal, 2 for above normal and 3 well above normal), glucose (1 for normal, 2 for above normal and 3 well above normal), smoking (binary), alcohol intake (binary), physical activity (binary), Presence or absence of cardiovascular disease (binary).

3.1 Data preprocessing

Prior to training our machine learning models, the proposed data was cleaned in order to eliminate:
- Records with missing one or more parameter.
- Records with suspicious or flawed values.
In this initial study, and in order to simplify the testing step, we focused our interest on the following features as the models inputs: Age, height, weight, arterial pressure high, arterial pressure low. In addition to this, we appended an output feature to the dataset following the study of T. M. Trana and N. M. Giang on paper [25] and the American heart association (heart.org) that categorizes severity case of patient based on their blood pressure levels (table 1):

Table 1. Output feature categories

<table>
<thead>
<tr>
<th>Output feature</th>
<th>Value</th>
<th>Blood pressure interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>-1</td>
<td>&lt;120</td>
</tr>
<tr>
<td>Elevated</td>
<td>0</td>
<td>120 - 129</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>1</td>
<td>130 - 139</td>
</tr>
<tr>
<td>(Hypertension stage 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High blood pressure</td>
<td>2</td>
<td>140 or Higher</td>
</tr>
<tr>
<td>(Hypertension stage 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertensive crisis</td>
<td>3</td>
<td>Higher than 180</td>
</tr>
</tbody>
</table>

The final resulting dataset was divided into 70% for training and 30% for testing. Afterwards the dataset was normalized using the z-score method in pursuance of creating a comprehensive dataset that matches the various numerical elements to the output:

\[
\text{score} = \frac{\text{value} - \mu}{\sigma}
\]

Where \(\mu\) is the mean value of the parameter and \(\sigma\) is the standard deviation of the parameter.

3.2 Implementation technologies and methods

In the direction of comparing the classification algorithms mentioned before, we opted for the following technologies:

- Python: An interpreted, multi-paradigm, cross-platform programming language.
- Scikit-learn: A free Python library for machine learning.

Python and Scikit-learn library come with an extensive repository for machine learning algorithms that are ready for implementation. The training and testing of the classification algorithms were carried out on a platform with the following configuration:

- Processor: Intel i5-5200U CPU, 2.20 GHz.
- Memory: 4 GB.
- Operating System: CentOS 7.

3.3 Results and Analysis

The performance of each of the machine learning classification algorithms selected for comparison was measured using the following metrics:

- Accuracy Classification Score: This metric corresponds to the number of correct prediction (percentage) based on the dataset (70% training and 30% testing).
- Training Time (Latency): The time necessary to carried out the training operation.
- Precision Score: this score represents the validity of the results (Mathematically, true positives divided by all positives).
- Recall Score: refer to the level of completeness of the results (Mathematically, true positives divided by relevant elements).

Following are the results:

Table 2. Machine learning classification algorithms accuracy and latency

<table>
<thead>
<tr>
<th>ML Classification Algorithm</th>
<th>Accuracy Score (%)</th>
<th>Training Time (in seconds)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>7.16</td>
<td>0.028</td>
<td>Linear regression predict a continuous value (linear) which explain the low score.</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>96.04</td>
<td>4.150</td>
<td>-</td>
</tr>
<tr>
<td>Naïve Bayesian networks</td>
<td>85.8</td>
<td>0.024</td>
<td>-</td>
</tr>
<tr>
<td>Multi-Layer perceptron</td>
<td>99.80</td>
<td>88.024</td>
<td>3 hidden layers with 10, 15 and 20 nodes respectively (500 iterations)</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>94.95</td>
<td>26.25</td>
<td>-</td>
</tr>
<tr>
<td>K Nearest Neighbor</td>
<td>84.41</td>
<td>0.875</td>
<td>Training was set with 5 initial nearest neighbors and “minkowski” for the calculation metric</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>99.59</td>
<td>0.068</td>
<td>Maximum depth of 5</td>
</tr>
</tbody>
</table>

Figure 4 illustrates the accuracy score of each ML classification algorithm used on this study where we can see clearly that “Linear regression” cannot help with the problematic in hand given the
low accuracy score that amounts to its core process that predict a continuous linear value based on the training parameters. The rest of the algorithms scores are very promising however, the “Multi-Layer perceptron” and “Decision Tree” algorithms scored the best accuracy in predicting the patient health condition given the input data. “Logistic Regression” and “Support Vector Machine” also presented a high accuracy score that highlight their ability as strong candidates for health condition predictions. Afterwards, “Naïve Bayesian” and “K Nearest Neighbor” presented an important accuracy score that features the significance of these algorithms classification capabilities.

Figure 5 displays the fastest ML classification algorithms of this investigation where we can see that “Naïve Bayesian” and “Decision Tree” are the quickest on the training time an thus making them the most accurate selection for a real time task. The “K Nearest Neighbor” shows also a remarkable output when it comes to the required training time.

Table 3 demonstrates the precision and recall score of the evaluated algorithms:

<table>
<thead>
<tr>
<th>ML Classification Algorithm</th>
<th>Precision Score (%)</th>
<th>Recall Score (%)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>-</td>
<td>-</td>
<td>Not applicable (constant linear value prediction)</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>95.07</td>
<td>96.14</td>
<td>-</td>
</tr>
<tr>
<td>Naïve Bayesian networks</td>
<td>87.67</td>
<td>86.36</td>
<td>-</td>
</tr>
<tr>
<td>Multi-Layer perceptron</td>
<td>99.74</td>
<td>99.75</td>
<td>-</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>94.29</td>
<td>95.13</td>
<td>-</td>
</tr>
<tr>
<td>K Nearest Neighbor</td>
<td>82.18</td>
<td>82.28</td>
<td>-</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>98.85</td>
<td>98.92</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 7 describes the precision and recall score of the ML algorithms involved in this study. Decision Tree produced the highest precision by eliminating most of the false positive and negatives. Multi-Layer perceptron, SVM and Logistic Regression revealed a high precision and recall scores, however these algorithms presented a poor outcome in regards to the latency (Figure 6). Naïve Bayesian and K Nearest Neighbor exposed promising results on precision and recall. Finally, linear regression cannot be compared to the other algorithms as it consists on predicting continuous linear values that will always be different from the output data in hand.

By analyzing these results and through the combination of the accuracy, precision and recall scores and also the necessary training time of the ML classification algorithms, one can clearly distinguished that “Decision Tree” is the ultimate ML classification model for the investigated issue (health condition) that requires a real time decision. “Logistic regression”, “Naïve Bayesian networks” and “K nearest neighbor” are some major contestant models that requires further investigations given their impressive outcomes during this study. Nevertheless, Multi-Layer Perceptron, SVM and Linear regression cannot be considered as solutions for the problematic in hand given the high latency of MLP and SVM and the low accuracy and precision of the linear regression algorithm.

As a summary, the proposed smart home design and technologies (section 2 and sub-sections 2.1, 2.2 and 2.3) in conjunction with the machine learning algorithm classification algorithms (sub-
section 2.4 and section 3), can deliver a powerful and intelligent environment capable of monitoring a patient health condition in real-time without the need of being admitted to a hospital or nursery. To the extent of our knowledge, rather than proposing singular solutions for every healthcare problematic, the proposed method and design would eventually lead to the construction of a smart healthcare ecosystem that reduces the pressure that come with the actual centralized healthcare systems and making it more distributed, case sensitive, more intelligent and ultimately comfortable.

4. CONCLUSION

In the final analysis, creating a smart home model for medical surveillance that incorporate trending technologies such as mobile application, sensors, wearables, cloud computing and machine learning capabilities could be a sustainable solution for healthcare. As a matter of fact, centralizing both the smart home and patient health condition management on the cloud would reduce the energy consumption on the patient (user) smart phone. In this design, the patient (user) smart phone become a bridge that connects and collects data from different medical sensors and wearables and consumes the cloud services required to monitor the patient health condition and route the necessary information to the practitioner. Consequently, the proposed smart home architecture exhibits an innovative model with many advantages for medical surveillance, thus can be a novel contribution to the smart healthcare paradigm. Additionally, machine-learning algorithms are an essential aspect of medical surveillance. Hence, exploiting the classification capacities of these algorithms on monitoring and predicting an arising health condition would drive the proposed smart home architecture to the artificial intelligence level. By means of testing and comparing different ML classification algorithms in this study, the “Decision Tree” algorithm was the foremost and the conclusive model for a real-time medical surveillance environment.

One remaining question is: how can machine learning improve the other smart home characteristics namely facial and gestures recognition, fall detection? From this question, we expect to navigate new investigations domains in the aim of enriching our research study.

REFERENCES:


Figure 2. The proposed smart phone mobile application design and architecture

Figure 3. The proposed Cloud main application panels

Figure 4. Machine Learning Classification accuracy
Figure 7. Machine Learning Classification precision and recall score