Elderly Healthcare System for Chronic Ailments using Machine Learning Techniques – a Review

R. L. Priya*, S. Vinila Jinny

Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Tamilnadu, India

Abstract

World statistics declare that aging has direct correlations with more and more health problems with comorbid conditions. As healthcare communities evolve with a massive amount of data at a faster pace, it is essential to predict, assist, and prevent diseases at the right time, especially for elders. Similarly, many researchers have discussed that elders suffer extensively due to chronic health conditions. This work was performed to review literature studies on prediction systems for various chronic illnesses of elderly people. Most of the reviewed papers proposed machine learning prediction models combined with, or without, other related intelligence techniques for chronic disease detection of elderly patients at an early stage to avoid emergency situations. This method provides a promising approach in the analysis of either structured or unstructured datasets to produce very substantial pattern discoveries. By defining the generic architecture for the prediction model, we reviewed various papers involved in similar fields, based on suggested methodologies and their associated outcomes. The study discussed the pros and cons of different prediction models using traditional and modern machine learning techniques.

Keywords: Machine Learning, Chronic diseases, Data mining, Internet of Things (IoT), Deep Learning.

1. Introduction

Although the value of health is universally agreed upon, its definition is not. Both the World Health Organization (WHO) and the United Nations (UN) [1, 2] define health in terms of well-being. They advocate a globally shared responsibility that all of us, i.e., states, international organizations, pharmaceutical corporations, civil society, and individuals, bear on the health of the world's population. Moreover, data indicate that, in the year 2012, world population was composed of 11 percent (809 million) of ageing society and it is expected to increase by 22 percent (92 billion) by 2050. According to the WHO statistics and Elderly Status Report published by the Government of India [3], aging is directly correlated to a number of new health problems which challenge the current healthcare services. The dependency ratio of ageing people also increased from 10.9 percent in the 1960s to 14.2 percent in 2011 [4].

Many researches [5, 8, 9, 47] have described the need to integrate social and health care services, which are essential in order to face the growing need of care in ageing societies. Most of the surveys state that a significant percentage of ageing population [49] have greater levels of acute and chronic morbidities. According to a survey report of Knowledge Base on Population Ageing in India (BKPAI) [6], nearly 65% of the total elderly population are suffering from at least one chronic morbidity. The most commonly reported chronic morbidities among elder community are diabetes, hypertension, asthma, heart disease, arthritis etc. Recent survey report [7, 10] reveals that there are mainly four Non-Communicable Diseases (NCD), which are Cardio-vascular disease (CVD), cancer, chronic respiratory diseases, and diabetes, which are high in mortality rate. Also, these reports indicate that

---

*Email: priya.rl@ves.ac.in

elders living in urban areas are having higher prevalence of lifestyle-associated diseases, such as hypertension, diabetes and heart diseases.

The broad aim of this study is to aid the average life expectancy of the ageing population and ensure timely healthcare by using current cutting-edge technology innovations. The survey is conducted to discuss the different Machine Learning (ML) prediction models, which will help to support the elderly population in every country.

This review of literature started in August 2015, with the search included looking for keyword indices on various reputed journal publishers, such as Springer, ScienceDirect, IEEE Xplore, IJCA, Research Gate, and other online databases, for articles that include phrases entitled "Healthcare and ML". During the period from 2010 to 2019, 2545 articles were published. After topic filtering, this number is reduced to 114 articles that relate to the specified keywords. Out of these 114 articles, only 55 articles are used in this review.

The remainder of the paper is organized as follows: An overview of ML is discussed in section 2.0. Section 3.0 describes the various ML algorithms and provides a comparative study of related approaches used in different elderly health care systems.

2. ML overview

ML is more about a sub-symbolic process for making computer systems to adapt their actions in an accurate manner [11]. This was the basis of artificial intelligence. Most familiar areas of ML research in the healthcare industry involve neural networks, deep learning and data mining. In supervised learning, the system receives the correct response for each input. The unsupervised learning means that the system learns the significant patterns based only on the given inputs. Learning in biological systems always begins with unsupervised learning. In the reinforcement model, the system learns with limited feedback as well. Evolutionary computing is a rapidly growing area of artificial intelligence.

2.1 ML Approaches

ML approaches are generally classified into four models such as supervised learning models, unsupervised learning models, and reinforcement learning models. With over a hundred ML algorithms available, choosing the right algorithm is a challenging process. There is no way to determine if a particular ML algorithm is the optimal choice for all problems.

Based on review articles, 90% of existing healthcare systems use supervised learning algorithms and hardly 10% apply either unsupervised or reinforcement learning algorithms. It is evident there is wide application of structured disease prediction or diagnosis systems in the healthcare sector.

2.2 ML algorithms

From the reviewed articles, a variety of ML algorithms [12] are used. In the case of supervised learning, different classification and regression algorithms are applied. Mostly, algorithms like Support Vector Machines (SVM), Naive Bayes (NB), Artificial Neural Networks (ANN), K-Nearest Neighbor (K-NN), Ensemble method, Logistic Regression, and Linear Regression (LR) are used in the existing systems. A number of unsupervised learning algorithms as defined in existing health prediction systems [13,14] like K-Means, Hidden Markov Models (HMM) and Market basket analysis are employed in unsupervised models. Many ML algorithms are utilized in the reinforcement learning model [14] such as Q-learning, Temporal Difference (TD) learning, Generative Adversarial Networks (GAN), etc.

According to the review articles, a majority of the research papers in healthcare communities use regression algorithms, 33% use classification algorithms, and 14% use clustering-based algorithms as one of the traditional ML algorithms. It also shows that among various ML algorithms, such as regression and classification, are the most widely utilized for prediction in the geriatric healthcare area.

3. ML techniques in elderly health care for chronic ailments

In most of the countries, particularly in developing countries, people customarily visit various specialists in different clinics. In such cases, it is difficult to combine their medical record results and conclusions. Secondly, collection of personal medical records, especially of elderly people living in rural places, is another major issue. In order to build a medical database, the health data are primarily pooled from various sources, like clinical data, claiming and cost data, pharmaceutical data, and behavior and sentiment data [16]. In such cases, ML techniques are mainly used in assisting the healthcare professionals for diagnosis and disease detection, which then leads them to make more definite clinical decisions at the right time [14].

3.1 Major chronic ailments among elders

The most important challenging burden of the healthcare system is to detect, monitor and manage the chronic ailments [48] and social behavioral disorders [15] at an older age. As people grow older, it becomes essential to determine the frequency of occurrence and patterns of chronic diseases, as well as the multi-morbidity cases among them [17,19, 20, 21].

According to the BKPAI report, the major types of acute problems are fever, hypertension, diarrhea, asthma, gastritis, and arthritis. Chronic diseases are being considered as a major contributor to many diseases. They include, for example, hypertension, diabetes, asthma, arthritis, Parkinson disease, dementia and heart diseases. One of the major factors for the increasing need of nursing care in elders is due to their loss of Activities of Daily Living (ADL), as well as chronic diseases [18]. The BKPAI survey [6] indicates that among various chronic morbidities, 29 percent of elderly are suffering from arthritis. The prevalence rate of arthritis is at a higher level among rural elderly women. It may increase with the increase in age and decrease in education years. Disability of elders greatly depends on the context of arthritis. It can be immobilizing, and it comes in many forms.

Due to the increase of urbanization and socio-economic differences, hypertension is known as a lifestyle disease. Previous surveys [22, 27] showed that almost 21 percent of elderly are reported with hypertension. The prevalence rate of hypertension is at a higher level among elderly women in comparison with elderly men. Research indicates that the hallmark of diabetes is having high blood glucose levels. It is a group of diseases that affects the body's ability to produce or use insulin correctly [24]. Ten percent of the elderly are reported with diabetes in a recent survey, conducted using bivariate and multivariate analysis [23]. Diabetes is defined as one of the lifestyle diseases as it is more prevalent among urban elders and its rate increases with the increase of education and wealth. Studies proved that there is a strong relationship between comorbidity factors (such as hypertension, diabetic, cardiovascular disease etc.) and obesity in elders [32]. Obesity is considered to be an important factor for chronic inflammation due to lack of physical activity, increase in intra-abdominal fat, non- restriction in diet, etc.

Surveys showed that around eight percent of elderly are having asthma. The prevalence of asthma is at a higher rate in elderly men living in urban areas as compared to elderly women and rural areas. Many studies showed that stress can also worsen asthma [25, 26]. Heart diseases include heart attack, stroke, cardiac arrest, high blood pressure, peripheral artery disease, cardiovascular condition, and heart failure. Among chronic diseases, heart attack (acute myocardial infarction) is one of the dangerous diseases faced by elderly patients [20]. A more recent study found that the risk of cardiovascular disease can be reduced by exposure to blue light, which decreases blood pressure of patients as well [28]. Emotional stress is considered to be one of the major causes of certain types of heart disease.

Parkinson's disease (PD) occurs due to a neurodegenerative disorder which affects nerve cells in the part of the brain that controls muscle movement. Nearly one percent of elderly people are affected by PD and its prevalence rate increases with age [29]. The symptoms have a greater impact on seniors and contribute to higher mortality. Alzheimer's disease is the most common form of dementia, affecting more than 5 million people worldwide. There are other types of dementias, like memory loss and impaired cognitive function that impact seniors badly [30, 31]. The prevalence rate of Alzheimer's and dementia are higher among rural elderly women, as stated by the BKPAI survey reported in 2011 [6].

3.2 Generic architecture of ML prediction model

As reported by the survey carried out by National Center for Biotechnology Information (NCBI) bookshelf, old aged people are more likely to suffer from chronic diseases, like CVD and circulatory disease cancers [19, 42], rather than acute problems. The cardiovascular morbidities and obesity [32, 33, 46] are more associated with certain risk factors, like aging, hypertension, smoking, diabetes, cancer etc. Tummala *et al*. [33] states that it is a challenging factor to diagnose cardiovascular morbidity. According to the existing social structure and culture, depression is another major factor for chronic diseases affecting the elderly population. For chronic diseases, as discussed in section 3.1, hundreds of ML approaches using different types of data and different assumptions are developed in senior healthcare systems, which involve disease prognosis, diagnosis, treatment suggestion and patient care [34, 35]. Currently, the healthcare system is undergoing a shift from a traditional ML approach to a modernized patient centered approach using the modern ML techniques.

The generic architecture of the ML prediction model is depicted in Figure-1. The generic diagram represents the working flow of the prediction system, consisting of major four components. They are data collection, data preprocessing, data implementation and data evaluation. Data collection is the extraction of different types of patient's data from various heterogeneous sources. Data used in most of the ML classification algorithms must be clean, consistent and complete. Hence, data preprocessing is employed before the dataset is trained well for better prediction. Once the trained model is developed, the test dataset is evaluated using various evaluation methods. Then the comparison is carried out in order to choose the best ML technique.
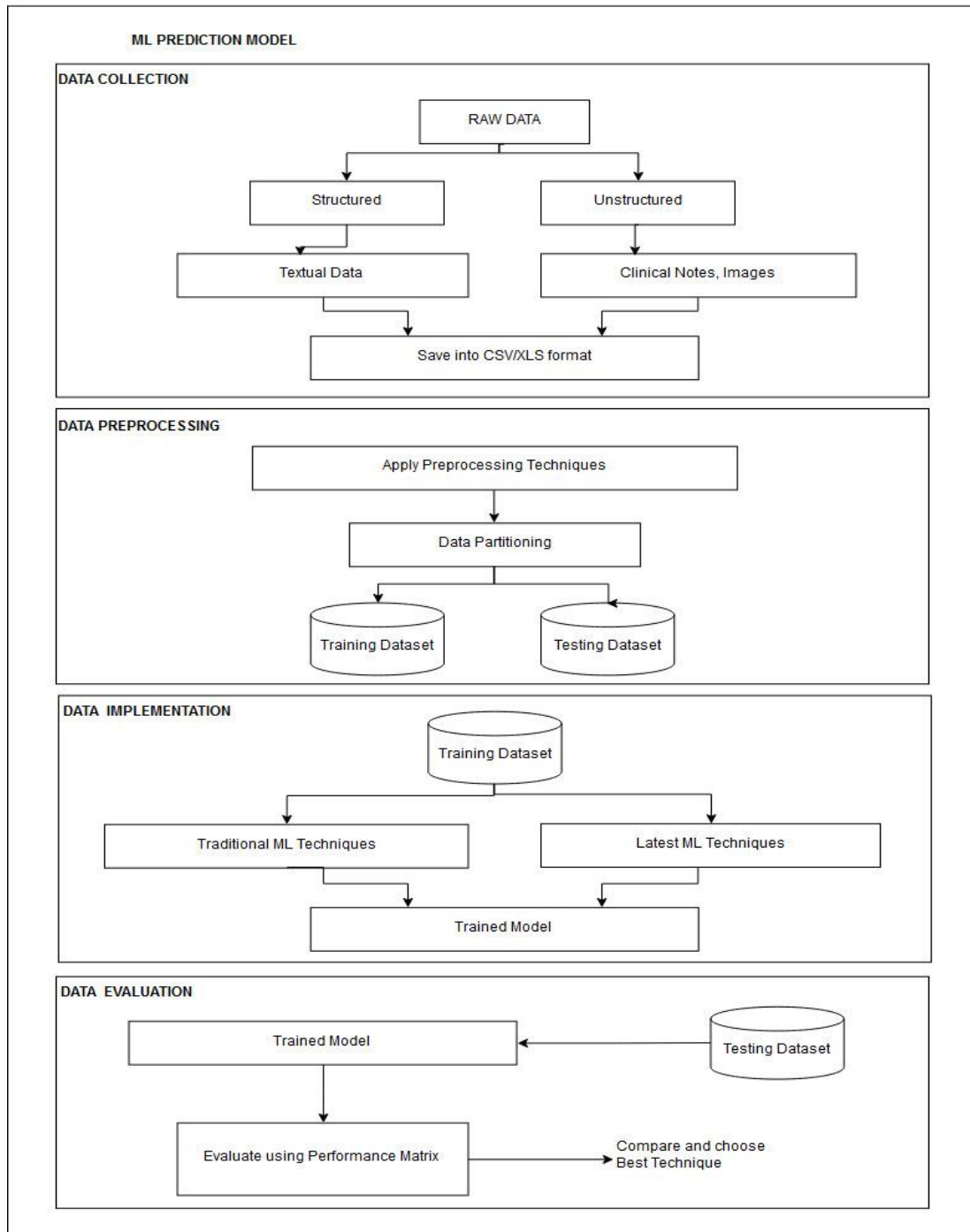

Figure 1- Generic architecture of ML prediction model

3.2.1 Data collection

Many existing ML based prediction models have revealed their importance to detect the key features, even from complex datasets. In the healthcare domain, patient's data are mainly available either in the form of structured or unstructured format, as defined in Figure-2.
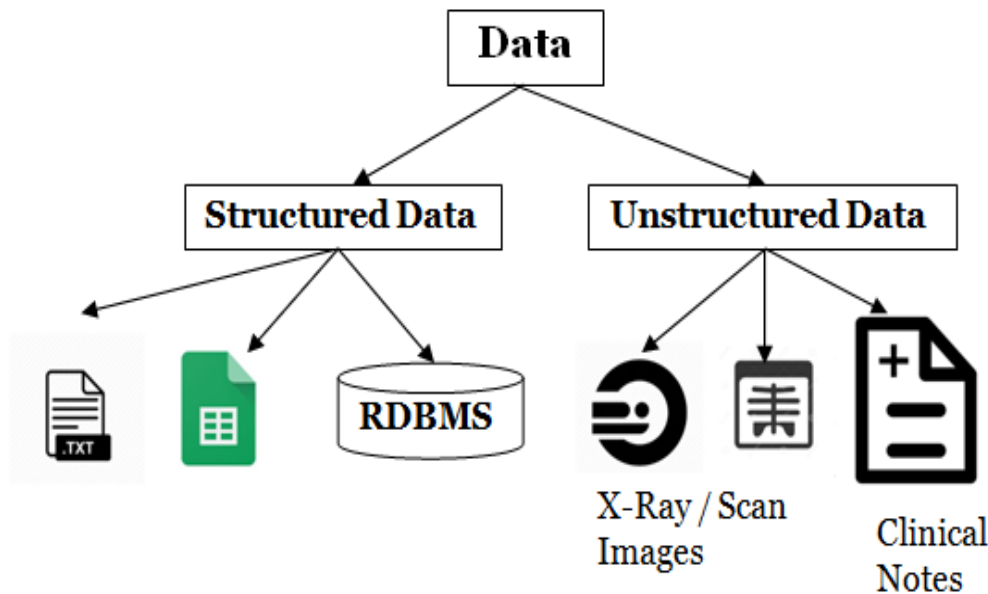
Figure 2-Classification of Patient's Data

The structured data consists mainly of two parts: patient's laboratory data and their demographic information, present in the form of tabular format. Unstructured data may include the elderly patient's narration about his/her illness, doctor's diagnosis and interrogation records, either in the form of images, textual information or clinical notes.

3.2.1.1 Structured Data

Data collection is an important process in every ML model. Based on the review, most of the prediction models use structured data as their input to the training set. For example, Chen *et al*. [13] proposed the prediction system to address the unique characteristics of regional chronic diseases using traditional ML mechanisms in case of structured data. It was an effective prediction model that used both structured and unstructured data over real-time hospital data. But the accuracy is restricted on the diversity features provided by clinical data. Ishita et al. [31] defined a predictive system, where data was acquired by interviewing nearly sixty elders with the help of Geriatric Depression Scale (GPS). The model is trained using various socio-demographic parameters such as age, gender, income status, education, employment, family type, etc., along with other comorbid conditions. Here, the researcher performed a comparison study with five different ML algorithms to choose the best model, which was found to be Bayes Net, for the prediction of depression among seniors. But the accuracy of the model was satisfactory and it can be further improved by considering more factors or applying better algorithms with optimization. The study [12] was conducted to identify the various factors, such as monthly income, diagnosis, depression, discomfort and perceived health status, which dictate Health-Related Quality of Life (HRQoL) of the elders with chronic diseases. A prediction model [12,13] was designed using various traditional ML models by considering 716 cases obtained from a Korean structured database system. The advantage of such a system is the identification of related influencing factors that affect HRQoL of seniors. It had a greater effect on building a better prediction system using ML techniques. But it failed to examine the impacts of one variable over the other, and the nature of influencing factors are also unknown.

Similarly, there are many intelligent electronic devices which are used to capture data and perform analysis of chronic disease management, as the patients go about with their daily routine. Wearable devices with cloud storage [50,51] and data analytics are also used to indicate fall of any one or two specific medical issues related to chronic disease [44]. For example, wearable devices can indicate heart condition and initiate an emergency call in the case that a heart attack is detected [36,37].

Another example is the bluetooth emergency system [16] which is devised for the wireless detection of heart attack using a wrist watch. Further, the telecardiology system [16] is used for heart attack detection and adds benefits to the rurally located patients. Danielsen et al. [36] proposed the fall prevention tool especially for elderly people using the combined technology of wearables and ambient intelligence [43].

Hussain *et al*. [34] proposed an Internet of Things (IoT) enabled framework and studied its challenges in data acquiring. It uses Wireless Body Area Networks (WBAN) and various health sensing devices and smart phones as the health monitoring devices. The above proposed intelligent frameworks are well designed to handle emergency cases by providing real-time medical services in smart cities. But it is a general model that does not provide proper support to rural and semi-urban areas of the country. In the recent era, IoT became a blooming technology, especially in the healthcare industry, which offers smart intelligent services. The connectivity in IoT enables communication among a variety of objects in a faster mechanism, being linked to the local and remote networks [16]. Thus, smart environments and wearable devices [37] are found to be more appealing in the health industry for the patients of all ages. The recent market studies state that there are numerous wrist watches / bands that are available in the market. Few fit bands allow users to maintain a healthy lifestyle by measuring a person's day-to-day activities. Certain bands may monitor only specific physical fall incidents or health discomforts of an individual.

3.2.1.2 Unstructured Data

Many more existing promising approaches used for data extraction of unstructured format are applied in the healthcare field. Diversity of climate and living habits of a certain region may become the prominent causes of regional chronic diseases [38]. For example, Chen *et al*. [13] proposed a model to predict regional chronic diseases using modern ML techniques, such as Convolution Neural Network based Multimodal Disease Risk Prediction (CNN-MDRP) for unstructured data of images. These models are best suited to handle the incomplete hospital data. They lack the accuracy of the model as they greatly depend on the features of hospital data.

A recent study [16] discussed the development of a customized healthcare system for patients to manage chronic diseases in an effective method. Such a system is developed using open-source big data Hadoop with text mining. Such technique helps in the transformation of unstructured data into structured data. But the biggest challenging task here is handling unstructured dataset which includes grammar mistakes, spelling mistakes, short phrases, etc., obtained from patient lab reports and vital signs. Such a system helps in the improvement of health care systems. It is very important to have proper verification and testing techniques to deal with the different critical domains.

Pradeebha *et al*. and Bhuvaneswari *et al*. [26, 39] addressed the feature extraction of unstructured data, in the form of images using various filtering techniques, such as the Gabor Kernel Filter (GKF), Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). Bhuvaneswari *et al*. [39] described the feature selection, which was carried out by various methods like information gain, Correlation-based Feature Selection (CFS), Principal Component Analysis (PCA), Genetic Algorithm (GA) and fuzzy c-means clustering. Among these methods, CFS has the maximum accuracy.

The feature extraction work can be further improved by other variants of LBP in order to detect a greater number of lung diseases. Li *et al*. [27] discussed the handling of structured, semi-structured and unstructured health datasets for detecting hypertension patients. They used Natural Language Processing (NLP) and text-mining techniques for the conversion of unstructured data into a structured format.

Zia Uddin *et al*. [37] proposed a multimodal human activity disease prediction system using multiple wearable sensors. The obtained sensor data are then encoded to model time-sequential information using Recurrent Neural Network (RNN) algorithm. But RNN is limited to process long-term information, which can be resolved using Long Short-Term Memory (LSTM). Xue *et al*. [40] also stated that LSTMs are used to perform many sequence learning tasks. Purushotham *et al*. [41] demonstrated that deep learning models perform better on different healthcare prediction models. It is best suited to handle large volumes of datasets and high complexity levels. It provides benchmarking evaluation results for deep learning and ML models by using huge records of raw clinical time series data as input features to the prediction system. Latest research [45] proved that deep learning methods provide better performance in clinical intervention prediction tasks using Intensive Care Units (ICU) data sources. The researchers demonstrated the benefits of feature selection using the LSTM method.

But the experiment was carried out only in a forward-facing manner to perform hourly prediction of both onset and weaning interventions.

3.2.2 Data preprocessing

Collection of raw data from different sources usually tends to be incomplete, inconsistent, missing values and noisy. Analyzing such data may produce misleading results. Hence, the quality of data and its representation are the foremost criteria before the execution of the analysis process. Data pre-processing is considered to be an important step in data mining and ML models. It may include data cleaning, normalization, transformation, feature selection and extraction. The output of data pre-processing is the input to the training set of ML models.

The major steps in data preprocessing are namely data cleaning, data integration, data transformation and data reduction. Data cleaning is used to identify outliers in order to smooth out noisy data, correct inconsistencies present in data, and fill with appropriate data in missing values. There are various techniques to fill out the missing values, either using global constant labels such as "Unknown", "Not Applicable" or using measures such as mean, median and central tendency for an attribute. Noisy data can be handled by different binning techniques, such as smoothing by bin means, bin medians, bin boundaries, regression etc. Most of the data smoothing techniques are used for data discretization and data reduction methods. Normalization of data is of great use, especially for classification algorithms, such as ANN, K-NN, etc. Min-max normalization, z-score normalization, and normalization by decimal scaling are few other techniques used for data normalization. Data reduction or discretization uses certain techniques, such as binning, histogram analysis, clustering, decision tree etc.

3.2.3 Data implementation and evaluation

In order to predict, prevent and manage chronic diseases, an efficient and very minimal error prone health care management system is essential. Several data mining methods, like J48 decision tree, bagging and boosting algorithms were used and their performance was compared for three different age groups (including elderly group) in a chronic disease prediction model. Among these, the decision tree was considered as the most powerful classification technique in chronic disease prediction [23]. It was concluded that the performance of adaboost was better than that of decision tree and bagging. It can be further improved by the integration of ensemble methods with other traditional ML techniques, such as NB, SVM, base learners, and others.

Perveen *et al*. [23] constructed a classification model with better performance to classify diabetic patients. But this model does not necessarily yield better results for other disease datasets, such as heart attack, dementia, depression, etc. Zhang et al. [18] described a predictive model using the C5.0 method of decision tree to achieve higher accuracy using various health variables, like personal habits, social activities, mental state etc. The limitation of such a system is that it is restricted with the use of continuous samples only. It also fails to predict the multiple classes of disability states.

Another survey also stated that, instead of data mining methods for chronic disease prediction, it is better to apply hybrid models to enhance the performance and accuracy of the prediction system [22]. An automated predictive modeling was developed to support earlier diagnosis of depression using traditional ML techniques. Here, multiple ML classifiers, such as BayesNet, Multilayer Perceptron network (MLP), LR, decision tree and Sequential Minimal Optimization (SMO) were applied to train the system. The results clearly revealed that SMO provides better accuracy and high precision values, while BayesNet provides better Receiver Operating Characteristic (ROC) area and lower Root Mean Square Error (RMSE), as compared to the other ML classifiers. Thus, a performance comparative study helps in early diagnosis of depression among senior citizens and the ability to choose the best performance classifier [31]. The drawback of this model is the use of inaccurate feature selection for prediction. The influences of other equally important parameters may be considered for further improvement and wider acceptability of the system.

The learning model performance was measured by certain evaluation parameters, like accuracy, sensitivity, precision, recall, and F-score, among which Stepwise Logistic Regression (SLR) analysis model provided better performance with an accuracy rate of 93% and F-score of 49% [12,13]. This system failed to analyze the impacts of each prediction variable considered in ML techniques. Also, better measurement tools can be considered in order to address disease specific parameters and target specific population groups, hence achieving more accurate results.

The training dataset was learnt with three most widely used ML algorithms, which are NB, K-NN and Classification and Regression Tree (CART). A performance evaluation of the predictive model was

carried out by few measurement tools, like accuracy, precision, recall and F-score measures. CNN-MDRP could achieve better accuracy as compared to CNN-based Uni-modal Disease Risk Prediction (CNN-UDRP). The system achieved accuracy of 94.8% with faster convergence speed. It was observed that to attain a better trained model, measurements such as Area under Curve (AUC) and ROC values must be approximately equal to one [38]. The graph analysis demonstrated that when there is an increased rate of recall measure, it leads to lower risk of disease prediction. The paper has discussed deeply the handling of structured data using more efficient traditional ML algorithms and unstructured data using modern ML techniques, like CNN-MDRP, to yield better accuracy. Although the paper deliberates about the prediction of complex diseases, it fails to predict when the disease is at a high-risk rate. In such cases, predicting chronic diseases at an early stage for specific age groups will become a challenging task.

Recently, an algorithm was proposed to provide a proper guidance to recognize and diagnose heart failure at an early stage [20]. This plays a vital role in reducing heart failure rate as early diagnosis helps to initiate appropriate treatment processes by healthcare professionals. It was explained clearly about the evaluation algorithms for the diagnosis of patients with chronic heart failure. It did not discuss earlier detection of heart disease for any age group specifically. This can be achieved by applying an appropriate traditional ML algorithm with proper feature selection parameters.

The most challenging task in IoT enabled health monitoring systems is to easily handle emergency situations that occur due to chronic problems. It also helps to maximize the other potential benefits of superior services. Hussain *et al*. [34] proposed an IoT enabled smart environment and discussed the various challenges faced at data processing, management and emergency care. There are certain other examples of prediction models, where big data analytics can be merged well with ML algorithms, rather than any other technique [28]. The limitation discussed in this case is about the accuracy of the model. False negative results are one of the most dangerous predictions, in which the patient stays unnoticed until the disease reaches a critical situation. In such a model, it is better to use either transfer learning or deep learning algorithms, depending on the size of datasets.

Fernández-Caballero *et al*. [14] proposed a smart environment architecture for the detection of emotions in older people. The model uses SVM technique and a traditional ML algorithm for the geometric feature extraction, based on facial positions (58 facial points are considered). Different other ways also help to regulate the emotions, such as music performance and color variations in the environment. The study did not focus on any specific health care problem. Pradeebha *et al*. [26] employed SVM Radial Basis Function as classifier and obtained an accuracy of 98%. Li *et al*. [27] demonstrated the classification model using C4.5 and SVM methods. Such methods provide better accuracy due to ease of use and ability to handle various categories of data in the presence of an overfitting issue. The big challenge is to choose appropriate learning algorithms for the management of chronic disease prediction in elderly.

3.3 Comparative study of chronic disease prediction for elderly people using traditional and modern ML techniques

Based on the data collected from a few selected survey papers, this study examined various factors of chronic disease prediction models developed using traditional and modern ML techniques. A variety of ML algorithms have been used for chronic disease prediction. The comparative study is to provide researchers with the information needed to improve health statistics of older communities for choosing an appropriate ML algorithm. It discusses the research objective, data collection, feature selection methods, ML methods, evaluation scores along with the advantages and limitations of various studies.

The survey found that the majority of elderly population suffer from one or more chronic diseases, with a substantial number of elders having physical and functional disabilities. The most common input datasets considered for the development of chronic disease prediction models are the structured datasets. In certain studies [13] structured as well as unstructured dataset (Electronic Health Record (EHR), medical images and gene data) are considered, whereas few studies [37,44] used sensor data like Electrocardiogram (ECG), magnetometer, accelerometer, gyroscope and so on. It is observed that very few studies use feature selection methods for further processing of prediction models. Certain researchers used modern ML algorithms, such as Random Forest (RF) [12], CNN [13], and LSTM [45] as feature selection methods. Also, word embedding [13] and Synthetic Minority Oversampling Technique (SMOTE) [30] techniques were applied in choosing the best features of the prediction model.

Certain studies use statistical and ML methods for the selection of features to predict most of the chronic diseases. In such a proposed system, accuracy and performance levels are only at acceptable criteria. The study clearly indicates that very few researchers trained the model using modern ML algorithms. Chen et al. [13] discussed the approaches of CNN-UDRP and CNN-MDRP. Among these, the CNN-MDRP proved better by performance measurements, compared to the CNN-UDRP. In addition, RNN algorithm [37] was used for the training of the model, and obtained an accuracy of 99.69 %. Another study [30] made use of the CART algorithm for the purpose of training, achieving an accuracy of nearly 92%.

The most common traditional ML algorithms used in various studies are decision tree, SVM, NB and K-NN. One of the earlier studies [23] trained the model using ensemble methods, like bagging and adaboost algorithms, and attained an ROC score of 0.98%. These traditional ML techniques were found to be better for structured data. Such methods yield better accuracy due to simplicity, suitability for numerical as well as nominal attributes, even with the presence of an overfitting issue. Finally, the review specifies the advantages and disadvantages of every study of the survey work, as dictated in Table-1. The descriptive analysis determined that most of the modern ML techniques help to improve the performance of the health monitoring system.

The great challenge is to use an efficient smart cutting-edge technology with correctly chosen traditional ML or modern ML algorithms for the prediction of one or more chronic diseases based on real time and available health records. Moreover, health care prediction results are very much sensitive; they do not provide either true positive or true negative results. Furthermore, diverse smart environment techniques using any wearable device are likely to incorporate traditional ML and deep learning mechanisms, such as RNN, CNN-MDRP, CNN-UDRP for building more sophisticated prediction and feature extraction models. Also, to facilitate better computational speed, GPU is utilized along with smart edging technologies like IoT, Cloud-of-Things (CoT) etc.

Table-1 represents the summarization of prediction models for elderly healthcare community using traditional and modern ML techniques. The table is designed based on the data retrieved from selective survey papers.

Table 1: Prediction of Chronic Diseases for Elders using Traditional and Latest ML Techniques (The data collected from selected survey papers)

Authors & Article	Year	Data collection	Research Objective	Feature Selection	Traditional ML Methods	Latest ML Methods	Best Performance Evaluation Scores		Adv.	Limitations
							Accuracy	Others		
Soo-Kyoung Lee et al. [12]	2014	716 cases structured data	To identify HRQoL risk factors by applying in prediction model with chronic diseases	Reference & expert data normalization, Random Forest with optimal parameter weight	DT, RF, SVM, SLR	-	93	F-score: 0.49	Using identified influencing factors, helps in the improvement of healthcare interventions	failed to examine the impacts of one variable over the other and the nature of influencing factors are also unknown
M. Chen et al. [13]	2017	706 cases hospital structured & Unstructured data (EHR, Medical Image & Gene data)	To develop prediction model on a regional chronic disease of cerebral infarction	Word embedding (structured data) & CNN (unstructured data)	Naïve-Bayes, KNN, DT	CNN-based Unimodal Disease Risk Prediction, CNN-MDRP	Recall – 0.80 (NB) 94% (CNN-UDRP)	AUC – 0.4950 Recall – 0.98 (CNN-MDRP)	Research more focused on handling both structured & unstructured hospital data	Predicting high risk of patients is a difficult task with the combinations of structured and Text data.
Kumar, Pradhan & Singh [15]	2017	9852 patient's data	To understand the correlation between the chronic diseases and elderly people disabilities	-	Statistical analysis (Multi-variate analysis) was used	-	-	-	Findings have shown significant association between chronic diseases and disabilities.	It limits the scope due the cross-sectional nature of data.
P.O. Nwani, A.O. Isah [17]	2016	345 cases	To analyze the prevalence of chronic diseases and multi-morbidity among elders	-	-	-	-	-	The higher prevalence among chronic diseases and multi-morbidity are highlighted.	The researcher restricted with hospital admission data only and cannot be defined as generalized model.
Sajida Perveen et al. [23]	2016	4678 cases- CPCSSN dataset	To classify diabetic patients based on diabetic risk factors using DM techniques	-	J48 Decision Tree, Bagging, Adaboost	-	AROC – 0.98%	Chi-square value – 46.85	Ensembles techniques proven for better performance than conventional systems	More selective features can be applied to improve the performance of the system.

Author	Year	Dataset	Objective	SMOTE	Methods	CART, NN	Accuracy	F-measure	Advantages	Limitations
P. Chaurasia et al. [30]	2016	173 cases CCSMA Data	To deploy assistive enabled technology for elderly health-care.	SMOTE	DT, KNN, NB, SVM	CART, NN	92.48%	F-measure: 0.85	It helps to identify factors that influences technology adoption.	More chance of imbalanced data in real-data scenarios
Ishita Bhakta, Arkaprabha Sau [31]	2016	60 cases (UHTC)	To predict depression patients among seniors	-	Bayes Net, Logistic Regression, DT, MLP, SMO	-	95%	RMSE – 0.22	It helps physician to predict depression patients by considering socio demographic and other morbidity parameters.	Lack of feature selection techniques to produce better optimized model.
Md. Zia Uddin [37]	2015	Sensor data – MHEALTH Public dataset (ECG, magnetometer, accelerometer, and gyroscope)	To develop smart healthcare system for activity prediction	-	HMM and DBN	RNN	99.69%	-	The system proves to predict the underlying activity with high accuracy for real-time sensor data. Usage of GPU results in faster computational speed.	The limitation of this project only focused on activity prediction.
X.Q. Yu et al. [40]	2013	Dataset from WHO Data repository	To predict the effectiveness of different treatment methods on diabetic young and old patients	-	SVM	-	-	-	The project considers NCD risk factors to analyze the effectiveness of diabetic treatment in both older and young age groups.	Limited their analysis using only one ML method.
P. Verma, S.K. Sood [44]	2017	25 cases - UCI dataset and medical sensors (EEG and EPOC sensors)	To develop fog assisted platform for real-time health care services.	-	DT, KNN, NB, SVM	-	92.8%	Sensitivity- 90.4% F-measure: 96%	It is a patient centric approach for real-time healthcare monitoring system	Limited in versatility to define the diagnostic methods
Harini Suresh et al. [45]	2017	34,148 cases MIMIC-III (ICU and Hospital admission data)	To predict the clinical interventions and to determine interpretability of models.	LSTM	-	CNN	-	AUC-0.97	CNN models over multi-channel input proven a clinical trend for real patient trajectories.	Limited to forward-facing approach for predicting both onset and weaning interventions.

*Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database
*Cache County Study on Memory and Aging (CCSMA)
* Urban Health and Training Centre (UHTC)
* Synthetic Minority Over-Sampling Technique (SMOTE)
* Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-III v1.4)

Conclusions

According to the National Sample Survey Organization (NSSO), there has been a steady rise in elderly population. In rural areas, the health-related problems faced by the older citizens are more critical compared to those of other age groups. Moreover, very few health care centres exist in the villages, where health professionals may not visit such centres regularly. At the same time, there are still multiple technology related challenges that exist in such areas. Elderly patients will require more time, given the usually complicated nature of their histories, medications and co-morbidities, which also means that they require more testing. However, it is important for healthcare providers to recognize that older patients do not always follow what is in the textbooks; providers must be vigilant to ensure that they do not underestimate the potential burden of caring for the booming population of older individuals.

Since there is a severe scarcity of revolutions in the field of geriatrics, strong gradation in technology integration are required. The main objective of any healthcare system is to promote a healthy

environment, which enhances a better quality of life among older adults. In this regard, our work reviewed various research papers to achieve a comparison study of different prediction models. It is found that conventional and modern ML techniques play eminent roles in healthcare industry for the prediction of chronic diseases. The different methods of data collection, data preprocessing, ML techniques, evaluation models, and results obtained from the survey list are also discussed. Hence, the healthcare industry is urged to maintain a secured cloud storage with an appropriate learning prediction model, which helps the health professionals to diagnose, predict and control the occurrence of any chronic problems. We also conclude that integration of smart healthcare devices by applying traditional ML and modern ML methods, like deep learning models, could help to build more efficient and adequate prediction systems in the healthcare industry.

Compliance with Ethical Standards

- Funding: The above study was not funded by any organization.
- Conflict of Interest: The authors declare here that they don't have any conflict of interest.

References

1. World Health Organization (WHO). **2013**. "World Health Statistics report", URI: http:// apps. who.int/iris/bitstream/10665/81965/1/9789241564588_eng.pdf?ua=1. 2014.
2. World Health Organization. (**2012**). Born too soon: the global action report on preterm birth. World Health Organization. https://apps.who.int/iris/handle/10665/44864
3. Elderly Status : Govt. of India - Ministry of Statistics and Program Implementation Central Statistics Office [www.mospi.gov.in] http://mospi.nic.in/sites/ default/files/ publication _reports /ElderlyinIndia_ 2016.pdf
4. Situation analysis of The Elderly in India, Central Statistics Office, Ministry of Statistics & Programme Implementation, Government of India, June **2011**.
5. Caring for our elders : Early Responses, India Ageing Report - **2017**, published by United Nations Population fund, May 31, 2017
6. Building a Knowledge Base on Population Ageing in India (BKPAI), Report on the Status of Elderly in Select states of India - 2011, published in November **2012**.
7. World health statistics **2020**: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.
8. Elderly in India – Profile and Programmes **2016**, Ministry of Statistics and Programme Implementation, Government of India, url : [www.mospi.gov.in]
9. State of Elderly in India **2014**, Help age India, Fighting Isolation, poverty, neglect, ISBN: 9789384439354
10. Mathur S, Mathur N. **2015**. Health Status of Elderly in India – The Path Ahead, *Indian Journal of Community Health*. **27**(3): 298-303.
11. Stephen Marsland. **2015**. "*Machine Learning – An algorithmic Perspective*", Second Edition, CRC press, 2015.
12. Soo-Kyoung Lee et al. **2014**. Prediction Model for Health-Related Quality of Life of Elderly with Chronic Diseases using Machine Learning Techniques, Healthcare Informatics Research · April 2014 DOI: 10.4258/hir.2014.20.2.125
13. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang. **2017**. "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities," in IEEE Access, **5**: 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
14. A. Fernández-Caballero et al. **2016**. Smart environment architecture for emotion detection and regulation, *Journal of Biomedical Informatics*, **64**(2016): 55–73
15. Kumar, Pradhan & Singh. **2017**. Chronic Diseases and their Association with Disability among the Elderly in India, Social Science Spectrum, ISSN 2454-2806, **3**(1): 27-37
16. Alexander CA, Wang L. **2017**. Big Data Analytics in Heart Attack Prediction. J Nurs Care 6: 393. doi:10.4172/2167-1168.1000393 ISSN:2167-1168, **6**(2), 1000393.
17. P.O. Nwani, A.O. Isah. **2016**. Chronic diseases and multi morbidity among elderly patients admitted in the medical wards of a Nigerian tertiary hospital, *Journal of Clinical Gerontology & Geriatrics,* (2016): 83-86
18. Yuejin Zhang et al. **2016**. Study on Prediction of Activities of Daily Living of the Aged People Based on Longitudinal Data, *Procedia Computer Science*, **91**(2016): 470 – 477

19. Subhojit Dey, Devaki Nambiar, J. K. Lakshmi, Kabir Sheikh and K. Srinath Reddy. **2012**. Health of the Elderly in India : Challenges of Access and Affordability, Aging in Asia : Findings from New and Emerging Data initiatives, Washington (DC) : National Academics Press (US) ; 2012

20. S. Mishra et al. **2018**. Management protocols for chronic heart failure in India, *Indian Heart Journal*, **70**(2018): 105–127

21. Shailendra Kumar, Manas Ranjan Pradhan, Abhishek Kumar Singh. **2017**. Chronic Diseases and their Association with Disability among the Elderly in India, *Social Science Spectrum* ISSN 2454-2806, **3**(1): 27-37

22. Neesha Jothi et al. **2015**. Data Mining in Healthcare – A Review, *Procedia Computer Science*, **72** (2015): 306 – 313.

23. Sajida Perveen et al. **2016**. Performance Analysis of Data Mining Classification Techniques to Predict Diabetes, Procedia Computer Science, **82**(2016): 115 – 121.

I. Kavakiotis et al. **2017**. Machine Learning and Data Mining Methods in Diabetes Research, *Computational and Structural Biotechnology Journal*, **15**(2017): 104–116.

24. Charan Singh. Ageing Population in India: Select Economic Issues, Indian Institute of Management, Bangalore

25. R. Pradeebha, S. Karpagavalli. **2016**. Classification of Lung Disease using Local and Global Descriptors, *International Journal of Computer Applications* (0975 – 8887), **135**(6).

26. Dingkun Li et al. **2018.** Application of a Mobile Chronic Disease Health-Care System for Hypertension Based on Big Data Platforms, Research Article Hindawi, *Journal of Sensors*, Volume 2018, Article ID 3265281, 13 pages, DOI: [url] https://doi. org/ 10.1155 /2018/3265281

27. Rajat Mehta. Heart Disease Prediction Using Machine Learning and Big Data Stack, url: [https://dzone.com/articles/a-tutorial-on-using-the-big-data-stack-and-machine]

28. Chao Gao et al. **2018**. Model-based and Model-free Machine Learning techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's disease, Scientific Reports, published in 8 May 2018, DOI : https://doi.org/10.1038/s41598-018-24783-4

29. P. Chaurasia et al. **2016**. Modeling assistive technology adoption for people with dementia, *Journal of Biomedical Informatics*, **63**(2016): 235–248

30. Ishita Bhakta, Arkaprabha Sau. **2016**. Prediction of Depression among Senior Citizens using Machine Learning Classifiers, *International Journal of Computer Applications* (0975 – 8887) **144**(7), June 2016

31. S. Amarya et al. **2014**. Health consequences of obesity in the elderly – Review article *Journal of Clinical Gerontology & Geriatrics*, **5**(2014): 63 – 67 Published by Elsevier Taiwan LLC, Open access under CC BY-NC-ND license.

32. R. Tummala et al. **2017**. Utility of ankle-brachial index in screening for peripheral arterial disease in rural India: A cross-sectional study and review of literature *Indian Heart Journal* (2017), http://dx.doi.org/10.1016/j.ihj.2017.07.012

33. A. Hussain et al. **2015**. Health and emergency –care platform for the elderly and disabled people in the Smart City *The Journal of Systems and Software*, **110**(2015): 253–263

34. Guest Editorial. **2016**. Ambient intelligence for health environments *Journal of Biomedical Informatics*, **64**(2016): 207–210

A. Danielsen et al. **2016**. Increasing fall risk awareness using 1s: A fall risk awareness Protocol, *Journal of Biomedical Informatics,* **63**(2016): 184–194 http://dx.doi.org/10.1016/j. jbi.2016.08. 0161532- 0464 /_ 2016, published by Elsevier Inc.

35. Md. Zia Uddin. **2019**. A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system, *J. Parallel Distributed Comput*. **123**(2019): 46–53.

36. Lin Wang et al. **2017**. Disease Prediction by Machine Learning over Big Data from Healthcare Communities, DOI 10.1109/ACCESS.2017.2694446, IEEE Access

37. C. Bhuvaneswari, P. Aruna, D. Loganathan. **2013**. *International Journal of Engineering Research and Applications* (IJERA) ISSN: 2248-9622 www.ijera.com, **3**(4): 2517-2524

38. X.Q. Yu et al. **2015**. Prostate cancer prevalence in New South Wales Australia: A population-based study, *Cancer Epidemiology*, **39**(2015): 29–36, http://dx.doi.org/10.1016/ j.canep. 2014. 11. 009 1877-7821/ _ 2014, published by Elsevier Ltd.

39. S. Purushotham et al. **2018**. Benchmarking deep learning models on large healthcare datasets, *Journal of Biomedical Informatics*, **83**(2018): 112–134.

40. K. Kourou et al. **2015**. Machine learning applications in cancer prognosis and prediction, *Computational and Structural Biotechnology Journal*, **13**(2015: 8–17.

41. J. van Hoof et al. **2011**. Ageing-in-place with the use of ambient intelligence technology: Perspectives of older users, *International Journal of Medical Informatics*, **80**(2011): 310–331, 1386-5056 © 2011 Elsevier Ireland Ltd.

42. P. Verma, S.K. Sood. **2017**. Cloud-centric IoT based disease diagnosis healthcare framework, *J. Parallel Distributed Computing* (2017), https://doi.org/10.1016/j.jpdc.2017.11.018.

43. Harini Suresh et al. **2017**. Clinical Intervention Prediction and Understanding with Deep Neural Networks, Proceedings of Machine Learning for Healthcare 2017, JMLR W&C Track Volume 68

44. R. O'Caoimh et al. **2015**. Risk prediction in the community: A systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults Maturitas, **82**(2015: 3–21. http://dx.doi.org/10.1016/j.maturitas.2015.03.009 0378-5122/© 2015, published by Elsevier Ireland Ltd.

45. J.M. Cramm et al. **2015**. The Health of India's Elderly Population: A Comparative Assessment Using Subjective and Objective Health Outcomes, *Population Ageing* (2015), **8**: 245–259, DOI 10.1007/s12062-015-9122-2, published with open access at Springerlink.com

46. M. Evandrou et al. **2017**. Children's migration and chronic illness among older parents 'left behind' in China, *SSM - Population Health*, **3**(2017): 803–807, October 2017, 2352-8273/ © 2017, published by Elsevier Ltd.

47. M. Sultana et al. **2017**. Burden of chronic illness and associated disabilities in Bangladesh: Evidence from the Household Income and Expenditure Survey, *Chronic Diseases and Translational Medicine*, **3**(2017): 112 -122.

48. Impact of Cloud Computing on Healthcare Version 2.0, white paper article by Cloud standards customer council, February **2017**.

49. R. Hoque, G. Sorwar. **2017**.Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model, *International Journal of Medical Informatics*, **101** (2017): 75–84.

List of Figures

List of Tables