Article

DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering

Abhaya Kumar Sahoo [1,*], **Chittaranjan Pradhan** [1], **Rabindra Kumar Barik** [2] **and Harishchandra Dubey** [3,*]

[1] School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar 751024, India; chitaprakash@gmail.com

[2] School of Computer Application, KIIT Deemed to be University, Bhubaneswar 751024, India; rabindra.mnnit@gmail.com

[3] Center for Robust Speech Systems, The University of Texas at Dallas, Richardson, TX 75080, USA

* Correspondence: abhayakumarsahoo2012@gmail.com (A.K.S.); harishchandra.dubey@utdallas.edu (H.D.); Tel.: +1-469-618-2851 (H.D.)

check for updates

Abstract: In today's digital world healthcare is one core area of the medical domain. A healthcare system is required to analyze a large amount of patient data which helps to derive insights and assist the prediction of diseases. This system should be intelligent in order to predict a health condition by analyzing a patient's lifestyle, physical health records and social activities. The health recommender system (HRS) is becoming an important platform for healthcare services. In this context, health intelligent systems have become indispensable tools in decision making processes in the healthcare sector. Their main objective is to ensure the availability of the valuable information at the right time by ensuring information quality, trustworthiness, authentication and privacy concerns. As people use social networks to understand their health condition, so the health recommender system is very important to derive outcomes such as recommending diagnoses, health insurance, clinical pathway-based treatment methods and alternative medicines based on the patient's health profile. Recent research which targets the utilization of large volumes of medical data while combining multimodal data from disparate sources is discussed which reduces the workload and cost in health care. In the healthcare sector, big data analytics using recommender systems have an important role in terms of decision-making processes with respect to a patient's health. This paper gives a proposed intelligent HRS using Restricted Boltzmann Machine (RBM)-Convolutional Neural Network (CNN) deep learning method, which provides an insight into how big data analytics can be used for the implementation of an effective health recommender engine, and illustrates an opportunity for the health care industry to transition from a traditional scenario to a more personalized paradigm in a tele-health environment. By considering Root Square Mean Error (RSME) and Mean Absolute Error (MAE) values, the proposed deep learning method (RBM-CNN) presents fewer errors compared to other approaches.

Keywords: collaborative filtering; deep learning; CNN; health recommender system; MF; RBM; SVD

1. Introduction

Nowadays, everything is available through the internet. When people are going to buy any kind of product through the internet, they first search for any reviews or comments about that product. At that time people may become confused about whether that product is preferable or not based on comments. Thus, a recommender system provides a platform to recommend such a product which is valuable and acceptable for people. Such a system is based on the features of the item, patient preferences and brand

information. This filtering-based system collects a large amount of information dynamically from the patient's interests, ratings, choices or the item's behavior, then filters this information to provide more vital information [1,2]. The theme of data analytics and big data are not an unfamiliar concept. However, the way it is characterized is continuously varying. Various approaches are made to retrieve large quantities of data efficiently because there are a lot of unstructured and unprocessed data that need to be processed and can be used in various applications. Healthcare is the best illustration of the application of big data analytics in different spheres of influence [3]. Data and information are spread among healthcare centers, hospitals, clinics. Beside three Vs (volume, variety, velocity), the veracity of healthcare data is also important for its role towards improving healthcare. Veracity refers to the consistency and trustworthiness of data [4,5].

A recommender system has the capability to anticipate whether a person would purchase a product or not based on the patient's preferences. This system can be implemented based on a patient's profile or an item's profile. This paper explains the item based collaborative filtering-based health recommender system which provides valuable information to patients based on the item's profile. Nowadays there are many blogs and social forums accessible on the internet where people can provide opinions, reviews, blogs and different perspectives regarding products. After collecting ratings for any product by patients, the recommender system makes decisions about patients who don't give any ratings. A number of e-business websites are working with the support of a recommender system to increase their revenue in the competitive market [1,6]. Millions of patients buy their products through online e-commerce websites. After buying products, they give their opinions or any comments about that product in a respective web forum. Thus, generating revenue is the main goal of all entrepreneurs. Using this recommender system process, we can increase our sales productivity in the market. While the preferences made by customers can be described as being low-risk, choices made in other sectors may have more intense ramifications for the end patient. In particular, in the sector of healthcare, choices can be life-threatening as they are concerned with the life and safety of patients. The recommender system should not only support decision making and avert dangers or failures, but it should also monitor patients and dispense treatment as necessary, keep track of vital signs and communicate in real time via a centralized server in the context of healthcare. These functions address the suitability of HRS [7–10].

The rest of the paper is organized as follows. Section 2 presents an overview of a collaborative-based filtering recommender system. Section 3 describes related works. Section 4 presents the proposed RBM-CNN method. Section 5 shows a comparison of different approaches with the proposed method and experimental results. Section 6 presents the conclusions and opportunities for future work.

2. Background

2.1. Preliminaries and Basic Concepts of Recommender System

In recommender systems, two main entities play crucial roles, namely patients and products. Patients give their preferences about certain items and these preferences must be found out of the collected data. The collected data are represented as a utility matrix which provides the value of each patient-item pair that represents the degree of preferences of that patient for specific items. In this way, the recommender engines are classified into patient-based and item-based recommender engines. In a patient-based recommender system, patients give their choices and ratings of items [11]. We can recommend that item to the patient, which is not rated by that patient with the help of a patient-based recommender engine, considering the similarity among the patients. In an item-based recommender system, we use the similarity between items (not patients) to make predictions from patients. Data collection for recommender systems is the first job for prediction [12].
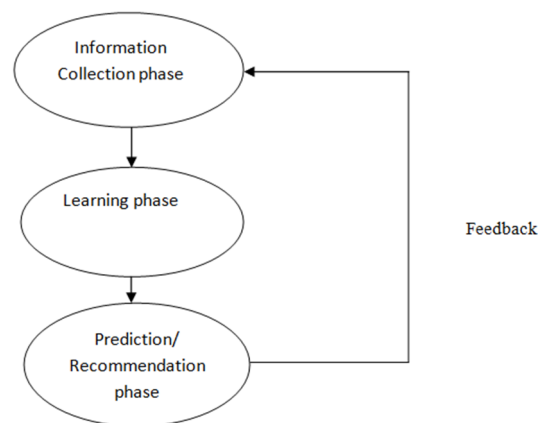
2.2. Phases of Recommender System

(1) Information Collection Phase: This phase collects vital information about patients and prepares a patient profile based on the patient's attributes, behaviors or resources accessed by the patient. Without constructing a well-defined patient profile, a recommender engine cannot work properly. A recommender system is based on inputs which are collected in different ways, such as explicit feedback, implicit feedback and hybrid feedback. Explicit feedback takes input given by patients according to their interest on an item whereas implicit feedback takes patient preferences indirectly through observing patient behavior [1].

(2) Learning Phase: This phase considers an assessment gathered in the former phase as input and processes this feedback by using a learning algorithm to exploit the patient's features as output [1,2,13].

(3) Prediction/Recommender Phase: Preferable items are recommended for patients in this phase. By analyzing feedback collected in information collection phase, a prediction can be made through the model, memory-based or observed activities of patients by the system [1,2].
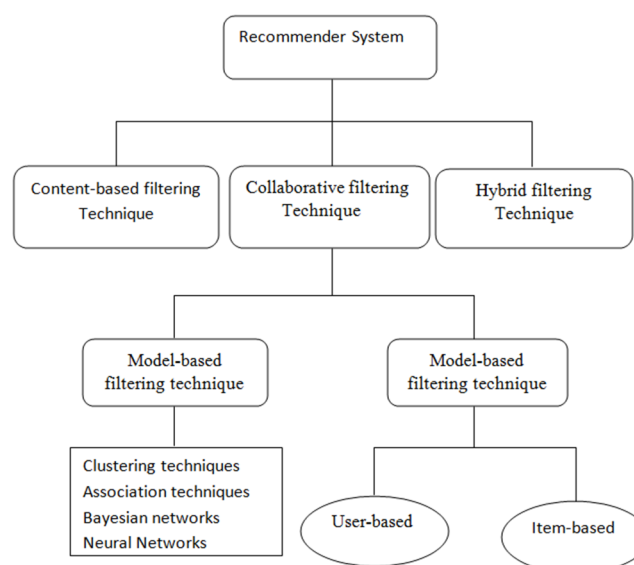
The phases of recommender system are represented in Figure 1.

Figure 1. Phases of the recommender system.

2.3. Different Types of Filtering Based Recommender System

There are three types of filtering based recommender system available, which is shown in Figure 2.

Figure 2. Hierarchy of Recommender System based on filtering.

(1) Content based Filtering Recommender System: The content-based filtering technique focuses on the evaluation of features and attributes of items to create predictions. Content-based filtering is normally used in case of document recommenders. In this technique, a recommendation is made based on patient profiles, which deal with the different attributes of items along with patient's previous buying history. Patients give their preferences in the form of ratings which are positive, negative or neutral in nature. In this technique, positive rated items are recommended to the patient [1,2].

(2) Collaborative based Filtering Recommender System: Collaborative filtering predicts unknown outcomes by creating a patient-item matrix of choices or preferences for items by patients. Similarities between patients' profiles are measured by matching the patient-item matrix with patients' preferences and interests. The neighborhood is made among groups of patients [14,15]. The patient who has not rated to specific items before, that patient gets recommenders to those items by considering positive ratings given by patients in his neighborhood. The CF in the recommender system can be used either for the prediction or recommender [16]. Prediction is a rating value $R_{i,j}$ of item *j* for patient *i*. This collaborative filtering technique is mainly categorized in two directions: memory based and model based collaborative filtering. Figure 3 explains the whole process of the collaborative filtering technique [4,17–22].
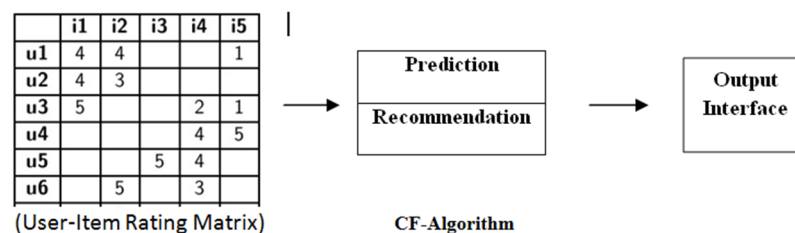
	i1	i2	i3	i4	i5
u1	4	4			1
u2	4	3			
u3	5			2	1
u4				4	5
u5			5	4	
u6		5		3	

(User-Item Rating Matrix)　　　CF-Algorithm

Figure 3. Collaborative Filtering Technique.

(3) Hybrid Filtering Recommender System: This technique comprises the above two methods in order to increase the accuracy and performance of a recommender system. The hybrid filtering technique is performed by any of the following ways: building a unified recommender system that combines both of the above two approaches; applying some collaborative filtering in a content-based approach, and utilizing some content-based filtering in the collaborative approach. This technique uses different hybrid methods such as the cascade hybrid, weighted hybrid, mixed hybrid and switching hybrid according to their operations [1,2].

3. Related Works

We studied the working principles of designing and developing a collaborative-based recommender system. Here we analyze the different health issues and develop an intelligent based health recommender system which provides high recommender quality to patients. There are different steps followed for developing and evaluating health recommender system using different machine learning algorithms.

3.1. Health Recommender System

With the rapid development of data mining and analytics, there has been a rise in the application of big data analytics in various domains. Healthcare systems have emerged as a promising sector in which big data analytics and its association have earned their own recognition and honor. The three main distinguishing characteristics of big data found in healthcare data are volume (the amount of data produced by organizations or individuals; the sources may be internal or external), velocity (the rate at which data is generated, captured and shared), and variety (the presence of data from a wide range of sources with different formats). Besides these three Vs, there is another V known as veracity (whether the obtained data is correct or consistent) which plays important role in healthcare. With huge

quantities of unprocessed data and information overload, recommender systems are becoming popular for their role in filtering large datasets and information. There is a need for a new HRS which can improve the health care system and handle the patients suffering from different diseases at a time. The HRS architecture is shown in Figure 4.
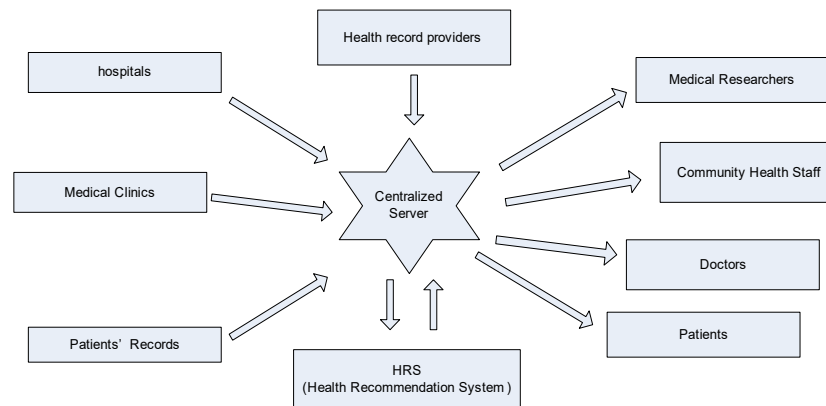
Figure 4. Health Recommender System (HRS) Architecture.

The recommender system is based on predictive analytics which predicts and recommends appropriate items to the patients. This system can be applied to specific applications. Healthcare analytics is a major area in big data analytics which can be incorporated into the recommender system. The health-based recommender system is a decision-making system which recommends proper healthcare information to both health professionals and patients as end patients. By using this system, patients are recommended the proper treatment of disease for avoiding a health risk, and health professionals benefit from the retrieval of valuable information for clinical guidelines along with delivery of high-quality health remedies for patients. This HRS should be trustworthy and reliable so that end patients can use this system to their benefit [23–25].

This HRS (shown in Figure 5) consists of different phases through which a particular item is recommended. These are training phase, patient profile process phase, sentimental analysis phase, privacy preservation phase and recommender phase. First, we have to collect a healthcare dataset to which we apply a feature selection and classification method. A major part of this HRS is to collect and prepare a profile health record (PHR) and patient database. PHR is a major concern as input for the recommender engine to predict and recommend health remedies to the patients. We collect useful information from patient database which is connected to PHR, for feature selection. Then, we apply a classification algorithm to classify and store the knowledge in a repository. The recommender process internally uses three sub-phases, namely the collection phase, the learning phase and recommender phase. After applying these phases using the patient database, proper treatments are recommended to patients and health professionals are recommended to valuable clinical guidelines and high-quality healthcare treatment. The sentiment analysis phase of HRS collects a patient's opinion for making the right decision in healthcare. This helps to discover the opinion of end patients with respect to a specific theme. The privacy preservation phase is used to provide privacy to the HRS so that valuable information is not changed. HRS is very helpful for decision-making in healthcare applications which can give offer significant value for society.
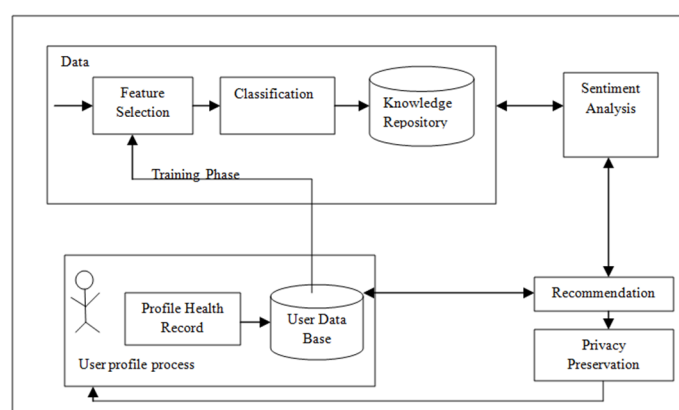
Figure 5. Health Recommender System (HRS).

3.2. Designing Health Recommender System

Many different methodologies are being developed in this newly emerging field. Here we proposed one methodology that is realistic and practical. In the first step, the software development team develops a problem statement which involves finding the objectives of the project. The problem statement is succeeded by a description of the project's importance. The designing team will perform a feasibility study of the project, including a technical assessment, cost estimation, and effort estimation. Once the problem statement is approved, the team can move to the next stage, the project development stage. Here, the team focuses on details of the projects. Because of the rise in project's cost in comparison to traditional ones, the team must do an economic feasibility study and explain why the project is cost-effective [26–28]. The project team should also provide background information on the problem domain as well as prior projects and research performed in this domain. After that, in Step 3, the design phase is implemented. The problem statement is broken down into a series of steps. Simultaneously, the independent and dependent variables or indicators are identified. The data sources are also identified; the data is collected, described, and transformed in preparation for data analytics. A very important step at this point is a selection of proper platform tools including Hadoop, Cloudera. In Step 4, the models and their findings are tested and validated and presented to stakeholders for action. Implementation is a staged approach with feedback loops built in at each stage to minimize the risk of failure. Different stages of the design methodology for HRS are described in Table 1.

Table 1. Outline of big data analytics in healthcare recommender system.

Steps	Description
Concept statement	Establish the need for big data analytics in healthcare based on the "4Vs".
Proposal	What is the problem being addressed? Why use a big data analytics approach? Background
Methodology	Objectives Variable selection and Data collection Data transformation Platform tool selection Analytic techniques, association, clustering, classification, neural network etc. Results
Deployment	Evaluation Testing

3.3. Framework for HRS

We need a framework for designing a HRS in cooperation with patients, doctors, surgeons and medical personnel. The architecture of the framework is divided into three parts (data collection, data transformation, data analysis and visualization). The first part is data collection. The data sources for the healthcare system have been categorized into (i) Structured data: organized data which has a predefined format, data type, and structure. Examples of such data include data generated from devices such as sensors, information about various diseases, their symptoms and diagnosis information, laboratory results, patient medical history, drug prescription, CT Scan, X-ray. (ii) Semi-structured data: data which does not conform to a data model but has some structure effective monitoring of patient's behavior. (iii) Unstructured data: data that has no defined structure, which may include medical prescriptions written in human languages, research notes, discharge summaries and so forth. Healthcare is a prime example of how the three Vs of data, velocity, variety, and volume, are an innate aspect of the data they produce. A large amount of data is spread among multiple healthcare systems, hospitals, health insurers, researchers, government institutions, etc. Different data sources are from prescription, clinical data, hospital records, patient information, vital signs, CT scans, X-rays and biometric fingerprints, physician prescriptions, etc. Healthcare automation systems constitute a branch of computational intelligence that applies reasoning methods and domain-specific knowledge to suggest recommendations like human experts. As with any other recommender domain, we must first understand the different categories of recommendations. The different categories are:

Nutritional data: Generating recommendations to augment nutrition. The doctor might change food habits so that patients get proper nutrition so that he/she can recover from illness or disease. Recommenders could be balanced food, substitution food items, less spicy meals, or additions to a diet.

- Physical exercise: Generating recommendations on what type of yoga and physical exercise the patients should do for quick recovery based on patients' requirements. The patient's requirements may include location, disease-related, weather, etc.
- Diagnosis: Generating recommendations on the diagnosis of patients by the doctor based on symptoms shown in similar cases.
- Therapy/Medication: Generating recommendations about different types of medication for a particular disease or patient-specific therapy.

The second part of the framework is the data analysis process. During the data analysis process, health-specific recommendations can be generated. We should first talk about patients who will be using this domain. The end-patient of the system is medical researchers, doctors, and patients. Apart from these end patients, there are other people who can benefit from the health recommender system (HRS) like pharmacists, clinicians, researchers. Minimizing the cost of healthcare should be the ultimate aim of these recommender systems. Analytical methods involve using Hadoop approach that uses MapReduce. This approach increases the speed of medical diagnosis and finding the optimal parameters for doctors so that he/she can detect the type of disease the patient is suffering and check the condition of the patient.

The third big part of the framework is the visualization part. This part contains elements that affect how recommended items should be presented. Visualization and knowledge representation techniques are used to present the mined knowledge to the end patients. The healthiest recommender is the one that should be chosen, but sometimes topic-specific criteria play a role in evaluating a product. Data-driven approaches apply data mining and machine learning methods to extract insights from the heterogeneous data. It provides individual recommenders based on the past learning experience and the patterns extracted from clinical data. A combination of information retrieval and machine learning can be used for the medical database classification. The entire framework of the health recommender system (HRS) is comprised of the following stages:

i.   Training Phase

ii.　　　Patient Profile Generation

iii.　　Sentiment analysis

iv.　　Recommender

v.　　　Privacy preservation

i. Training Phase

In order to detect various diseases like tuberculosis, cholera, flu, etc. doctors organize clinical tests on patients. Therefore, to study and analyze various diseases and find a cure for same, doctors require information through parameter and variables. Moreover, there has been a tremendous growth in the quantity of information being generated in healthcare. This phase includes data collection and accumulation. However, the absence of proper tools for the collection and accumulation of data will hamper the whole process. The whole process includes collecting various data and information of patients, demographic information of patients, diagnoses, research, clinical tests, patient's health record, real-time data from hospitals and clinic so that real-time data collection can enhance the effectiveness of the recommender.

ii. Patient Profile Generation

During this stage, for every patient, a patient profile is created which contains various information. For every patient, there will be a health record documenting the patient's clinical history. This record contains information from various sources, including the patient, doctors, hospitals, laboratory tests, CT Scan, X-ray, etc. If the new patient is admitted, then the whole process starts from the beginning, i.e., from the processing of data and the creation of a new patient health record. In the case of an existing patient, the system updates the record as per requirements.

iii. Sentiment Analysis

In order to support the patient-based recommender for the clinical services, it is imperative to make sure the patient trusts the whole system, i.e., system reliability to maintain privacy and confidentiality of patient data. Information with or without adequate medical data obtained from patients is personal and should not be misused.

iv. Recommender

From the extraction of rules and patient context, recommendations can be generated. Patients receive personalized recommendations. These recommendations can take the form of preventive and corrective measures, reasons for the causes of the disease, or a further process of treatment.

v. Privacy Preservation

The HRS requires the blending of various clinical information in order to enhance the recommender quality so that healthcare improves. Subsequently, ensuring the privacy of a patient's information plays a vital role in clinical research. In the proposed approach, the integrity of this information will be maintained while personal identity is effectively shielded [17,29–33].

3.4. Methods to Design HRS

We need a framework that is made up of different tools which satisfy the domain requirements and specific criteria of specific applications as shown in Figure 6. These tools present in the framework first put the focus on customer requirements and ensure that clients requirements are met first. The first tool vital for designing the framework is the use of participatory design. It refers to the active participation of stakeholders. Patients should play active patient while designing the framework because patient feedback can improve the whole system and remove lacuna present in the current system. Thus, feedback from the patient is essential. When patients draft the recommender system, they keep in mind health-related issues, not sales and marketing by pharmaceutical companies, because these systems

can become their personal assistant helping them to overcome health issues that are significant to them. The most demanding part is to hypothesize an existing framework in order to allow the large-scale participation of patients and doctors. These tools could help in the treatment without requiring the direct intervention of doctors.
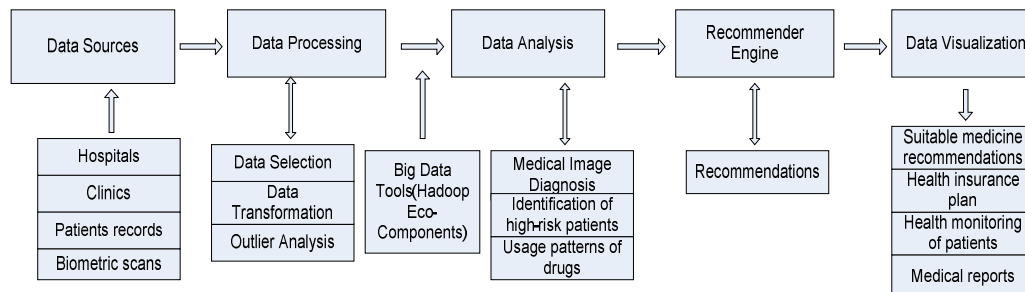
Figure 6. Health Recommender System Framework.

The second tool important to HRS, is the use of differential privacy. Differential privacy maintains data privacy and security which is main problem prevalent in a recommender system. Here, it is used for the sharing of a patient's medical history without revealing patient identities. So, privacy should be provided to the end patient. Patients are often unaware of privacy which presents a contradiction to their long-term interests. To implement an intelligent health recommender system, privacy must play a major role for the patients. The level of knowledge about privacy threats on the Internet is so important that different risk perceptions and levels of digital literacy are also related to technology. Patients are much more reluctant to share data in personal spaces.

The third tool to incorporate is adequate and proper communication. Communication is bidirectional (to the patient and to the recommender). Patients should be able to express in an undisturbed and hassle-free manner to doctors so that doctors can interpret the symptoms of patients and can give recommenders for a particular disease to patients. The visualization of data should address the purpose of the recommender system and able to understand the patients, doctors and their intentions. There should be a proper visualization tool in a recommender that fosters the patient's willingness to explore options and helps to explain individual recommendations. Since individual differences might play a vital role in the health sector, it is crucial to intensify research in this field.

Some common big data tools are used in health care sector, e.g., Data Cleaner, Apache Hadoop, Cassandra database etc. Apache Hadoop is the most prominent tool in the big data industry with its enormous capability of large-scale data processing. Hive is also one of eco-components of Hadoop which allows programmers to analyze large data sets on the Hadoop platform. It helps with querying and managing a large dataset. Data Cleaner is a data quality analysis platform which has strong data profiling engine. This tool is usually used for data cleaning, data transformation and data merging. Today's Cassandra database is widely used to provide an effective management of large amount of data. These tools are used to work with recommender engine in big data analytics [34,35].

3.5. Evaluation of HRS

For the success of the recommender system, it is very important to choose what type of criteria are used to evaluate the recommender system. Conventionally, recommender systems were evaluated based on criteria borrowed from information retrieval [9,31,36]. Common metrics used in the evaluation are:

i. Precision: The measure of retrieved instances that are relevant.
ii. Recall: The fraction of correctly recommended items that are also part of the collection of useful recommended items.

iii. F-Measure: It is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

iv. ROC-Curve: ROC Curve is a way to compare diagnostic tests. It is a plot of the true positive rate against the false positive rate. It is used to represent the relationship between sensitivity and specificity.

v. RSME: This measure defines the standard deviation of the residual errors, i.e., differences between predicted values and known values.

The evaluation criteria of the recommender system are very necessary to measure the strength of an HRS based on patient acceptance and satisfaction. By making system suitable for individual patients, the system can run as per patients' requirements so that patients will not face any problems, ultimately leading to better medical research. This includes patient diversity research, not just in regard to patient-specified results, but also in regard to the patient interface of a health recommender system. The pretentiousness of accuracy metrics and under-representation of metrics such as serendipity and coverage pose a serious challenge in a classic recommender system. Rare diseases are very uncommon but collection of data and case studies of similar cases can help a lot. Therefore, finding all relevant results is important for health recommender systems. Another very vital research issue is trust in recommender systems. If things take a worse turn, the doctor can program the system to take actions so that trust is maintained.

While designing health recommender systems, the person concerned should be careful and prepare plan according to requirements. The capability of a recommender system can be appraised in regard to the patient's external behavior. The measure of the effectiveness of a health recommender system depends upon behavioral evaluations. For example, when monitoring the health progress of patients and providing suggestions for treatment, keep track of activities. In the case of food restriction for patients, the system has a difficult time to measure its effectiveness, as some patients might smoke without informing the system. Some health recommenders may also aim at long-term behavioral changes and these must be tracked somehow as well. Once the treatment is administered, the system can continue monitoring the patient to determine if treatment is effective. The system should also take steps which can promote faster healing. We must consider those recommendations which do not cause any side effect because neglecting one health parameter can lead to another disease, e.g., changing food habits may lead to loss in body weight (a superficial health parameter), keeping our body fit but neglecting a balanced diet can hamper growth and metabolism. Before applying this approach for practical use, it must be ensured that systems are customer friendly and reliable. We must ensure too that the system delivers real time results.

4. Different Approaches Used in Health Recommender System

As data mining and recommender techniques are becoming popular, it has become important to find techniques that preserve data security and privacy. The personalized recommendations may be fruitful in attracting new patients, which raises a number of privacy matters. Based on survey results, new customers may look away from an e-commerce site because of privacy-related issues. Subsequently, while providing the recommender services, many inescapable security matters arise which penetrate the firewall through attempted unauthorized access. In order to shield against unauthorized data access and the misuse of personal information, it is imperative to conceal the patient database by secured methods while simultaneously creating a recommendation by making data more secure [29,37–39]. To sort out these problems, we will analyze different privacy-preserving based recommender systems, e.g., the matrix factorization model, SVD, etc. by which the system can generate recommendations without actually seeing the patient ratings.

4.1. Matrix Factorization

These are the most successful latent models which assist in solving high sparsity problems. In its primary form, the characteristics of both items and patients by an array of factors are derived from patient ratings patterns. These techniques have become favorable due to their better scalability and predictive accuracy. It is an influential technique to uncover the hidden structure behind data. It is used for processing large databases and providing scalability solutions. These are used in the field of information retrieval. These techniques map both patients and items to a joint latent factor space of dimensionality [30,40–43]. Each item can be represented by a vector q_i. Similarly, each patient can be represented by a vector p_u. These elements compute the extent to which product will be chosen by the patient. The dot product $q_i^T p_u$ represents the interaction between patient and item and is denoted by:

$$u_i = q^T_i p_u \tag{1}$$

4.2. Singular Value Decomposition

SVD is a well-known method for establishing latent factors in the scope of recommender systems to work on problems faced by the collaborating filtering technique. These techniques have become popular due to their good scalability and better predictive accuracy [44,45]. It is a popular recommender filtering technique that decomposes m × n matrix A into three matrices as A = USVT where U and V are two orthogonal matrices of size m × r and m× y, respectively; y is the rank of the matrix A. S is a diagonal matrix of size r × r. Singular values are present in the diagonal of the matrix [44]. It then decomposes S matrix to get a matrix Sk, k < r showing top k rated items (largest diagonal values), as shown in Figure 7.

Figure 7. Singular value decomposition.

The average value of patient ratings is utilized for filling up sparse locations in ratings matrix (A) to secure a fruitful latent relationship. Then, z-normalization of the matrix is done. Using SVD, the normalized matrix (Anorm) is divided into U, S, and V. Then the matrix Sk is acquired by taking only k largest singular values resulting in the reduction of dimensions of matrices U and V. After that, Uk, Sk and SkVT are determined. The final matrices can be used to find the prediction for the active patient. It can uncover the hidden matrix structure. This filtering is used mainly in CF recommender systems in order to detect the features of patient and item [35,46].

4.3. Variable Weighted BSVD (WBSD)

This type of collaborative filtering is introduced for solving problems in SVD based collaborative filtering and improve its accuracy and privacy. Here, we are using a variable weight so that we can change the weights as per our requirements [44,45]. If active patients are concerned about data privacy, they will disturb the data as per requirements. By disturbing data so that unknown patient cannot access data, privacy can be preserved. The patients receive many disturbing information and fail to classify items on basis of disturbing nature of data, and they cannot determine that which items are rated by particular individuals, and which are not [44]. A change in weight is computed using the formula:

$$\omega_i = e^{\delta_{max} - \delta_i} \tag{2}$$

where σ_i refers to disturb weight of patient i, σ_{max} represents the maximum disturb weight, and ω_i denotes the variable weight of patient i. The value of this variable lies between 0 and 1. The new Ak can be calculated as follows.

$$A_k = AV_kV^T{}_k \tag{3}$$

where the Ak represents the disturbed matrix, the row represents the data whereas column represents the item. This column behaves as the feature vectors according to the largest k eigen value of the matrix C_ω. Vk is an n*k matrix and $V^T{}_k$ is the transposition of matrix Vk. C_w is computed as follows:

$$C_\omega = \sum_{I=1}^{m} \omega_i x^T{}_i x_i \tag{4}$$

4.4. Deep Learning Method

Multilayer perceptron (MLP), auto-encoder (AE), convolutional neural network (CNN), recurrent neural network (RNN), restricted Boltzmann machine (RBM), neural autoregressive distribution estimation and adversarial networks (AN) are the main components of the deep learning method [10,33,47–49].

4.4.1. Multilayer Perceptron with Auto-Encoder

This is a feed forward neural network which has many hidden layers in which perceptron uses the arbitrary activation function. The auto-encoder (AE) is a model based on unsupervised learning which attempts to regenerate its input data in the output [50]. An auto-encoder neural network uses back propagation technique which calculates the gradient of the error function with respect to the neural network's weights, as shown in Figure 8.
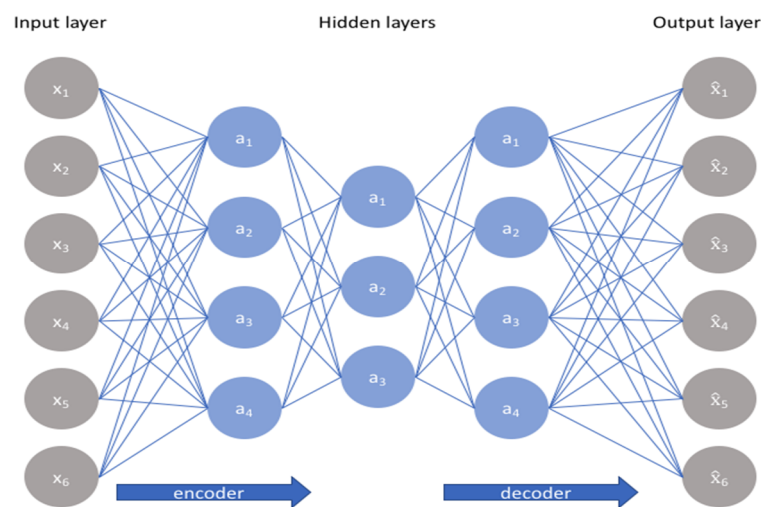

Figure 8. Auto-Encoder Neural Network.

Auto-encoders can be considered as heart of the representation learning. They encode input, usually by converting large vectors into smaller vectors that record the vectors' most significant features; which would be useful for data compression, data reconstruction for unsupervised learning and dimensionality reduction.

4.4.2. Convolutional Neural Network (CNN)

It is one kind of feed-forward neural network consisting of convolution layers which captures the global and local features for enhancing efficiency and accuracy. This network is used for modeling sequential data.

4.4.3. Restricted Boltzmann Machine (RBM)

The RBM is a generative stochastic artificial neural network that learns a probability distribution over a set of inputs. It is a two-layer neural network, which consists of visible layer and hidden layer [51,52]. There is no intra layer communication among the visible and hidden layer. It uses gradient descent approximation algorithms like contrastive divergence. This network is used for extracting features and attributes in a supervised learning algorithm. These models are energy-based models consisting of binary-valued hidden and visible units, as shown in Figure 9. The strength of connections between visible and hidden units Vi, and Hj, respectively is denoted by the matrix of weights W. It includes bias weights (offsets) for these units. Increasing the number of hidden units leads to better the performance of the model.
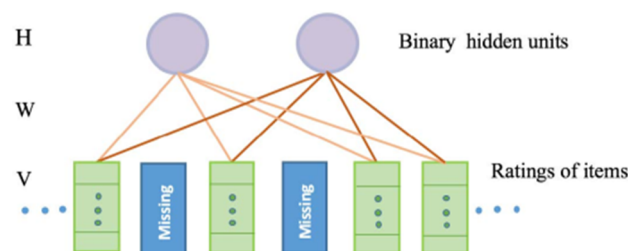

Figure 9. RBM Neural Network.

4.4.4. Adversarial Networks (AN)

This is a generative neural network, which consists of two parts-a discriminator and a generator. In a mini max game framework, the two neural networks are trained simultaneously by competing with each other.

4.4.5. Neural Autoregressive Distribution Estimation

This is another type of unsupervised learning based neural network consisting of the autoregressive model which is further complemented by a feed of forward neural networks.

5. Proposed RBM-CNN Based Health Recommender System

In the standard RBM all observed variables are related to all hidden variables by different parameters. Using an RBM to extract global features from full images for object detection is not promising considering how large the images are. To tackle this problem, there is a variant of the RBM model, called the convolution RBM (CRBM). The CRBM, similar to the RBM, is a two-layer model in which visible and hidden random variables are structured as matrices. Therefore, in this model, locality and neighborhood are definable both for hidden and visible units. The CRBM's visible matrix could represent an image and sub windows of it would denote image patches. The CRBM's hidden-visible connections are local and weights are shared among clusters of the hidden units. Therefore, it can be concluded that CRBM is better than RBM and CNN as it uses the features of both RBN and CNN.

i.   Load the healthcare dataset and also passheader=none since files don't contain any headers.
ii.   Load the ratings dataset
iii.   After that, rename our columns in these data frames so we can convey their data better.
iv.   Verify the changes done to the data frames.
v.   Data Correction and Formatting.
vi.   Merge no. of hospitals with ratings by hospital ID.
vii.   Display the result.
viii. Number of patients used for training.
ix.   Creating the training list.

    (a)   For each patient in the group for patientID.

    (b)   Create a temp that stores every health care's rating.

    (c)   For each health care in curPatient's health care list for num.

    (d)   Divide the rating by 5.

    (e)   Add the list of ratings into the training list.

    (f)   We will verify that we have finished adding in the number of patients for training and setting the model parameters.

x.   Train RBM with CNN 15 Epochs, with each epoch using 10 batches with size 100.

xi.   After training, the error is printed out by epoch size wise.

xii.   Select the input patient.

xiii.   Feeding in the patient and reconstructing the input.

xiv.   List the 20 most recommended hospitals for our mock patient by sorting it by their scores given by our model.

xv.   Find the mock patient's PatientID from the data.

xvi.   Find all hospitals the mockpatient has visited before.

xvii.   Merge all hospitals that our sample patients have visited with predicted scores based on his historical data.

xviii.   Merging hospitals.

xix.   Dropping unnecessary columns.

6. Experimental Result and Discussion

We conducted experiments on the data set of healthcare [37]. This healthcare dataset contains discrete ratings from 1 to 5 of 10,000 patients for 500 hospitals. This dataset is divided into training and test data in 75:25 ratios respectively. Here 10-fold cross-validation scheme is used while evaluating the results. We implement the proposed CRBM method using Tensor Flow and python. Our HRS with different approaches is designed and tested on a healthcare dataset which describes rating information along with details.

We need to choose parameter 'K' as no. of nearest neighbors. If K is very less, the data would lose the vital information, and if K is too big, it will lose its data privacy property. Therefore, parameter K should be selected properly. The approach of root mean absolute error (RMSE) is used here because it can be easily identified and measured so that we can compute the quality aspect of recommenders easily. It is utilized in this paper to exhibit the performance of the different techniques and their accuracy.

From Figure 10, it can be seen that the RMSE of the proposed RBM-CNN-based collaborative filtering method varies with variable K, and achieves a perfect value when K is equal to 10. The lesser the value is, the higher the accuracy is. This leads to better recommender quality.
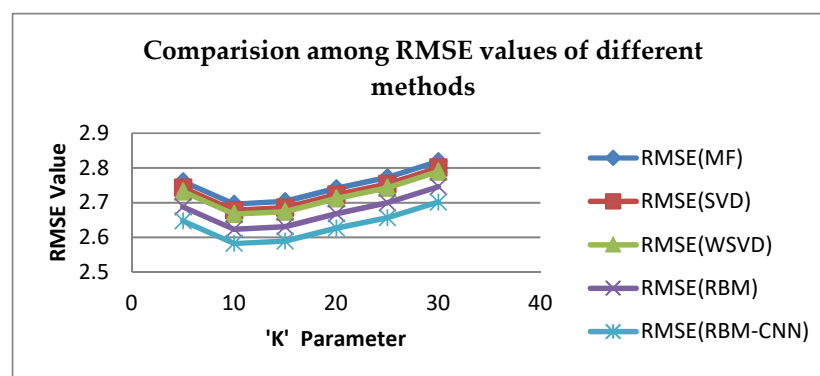

**Figure 10.** Comparison among K and RMSE of different methods.

Table 2 depicts for different K values, specifically, the results indicate that RBM-CNN is best in terms of RMSE for K=10.The experimental results indicate that RBM-CNN technique is best in terms of accuracy. As a result, RBM-CNN should be the preferred method of collaborative filtering. There are other metrics which can also measure recommender quality such as MAE (Mean Absolute Error), precision, Recall and ROC etc. An evaluation for each deep learning-based recommender system is tested with different number of epochs. Each method is evaluated with MAE as shown in Table 3. From Figure 11, it can be seen that the greater number of epochs, the higher the accuracy is. This leads to a better recommender quality.

Table 2. Comparison among K and RMSE of different methods

K	MF	SVD	WSVD	RBM	Proposed RBM-CNN
5	2.76137	2.74313	2.73219	2.68828	2.64707
10	2.69592	2.67776	2.66688	2.62337	2.58216
15	2.70339	2.6852	2.67413	2.63062	2.58931
20	2.74089	2.72274	2.71169	2.66818	2.62657
25	2.77257	2.75391	2.74288	2.69937	2.65616
30	2.81879	2.80047	2.78935	2.74584	2.70136

Table 3. Comparison among number of epochs and MAE of different methods.

No. of Epochs	MF	SVD	WSVD	RBM	Proposed RBM-CNN
0	0.14412752	0.13432652	0.12912452	0.12286516	0.11785
2	0.08904246	0.07924146	0.07403946	0.0677801	0.05876
4	0.07580589	0.06600489	0.06080289	0.05454353	0.04734
6	0.068728	0.058927	0.053725	0.04746564	0.03748
8	0.06597785	0.05617685	0.05097485	0.04471549	0.03471
10	0.06439402	0.05459302	0.04939102	0.04313166	0.03368
12	0.063298018	0.053497018	0.048295018	0.042035658	0.03206
14	0.062187785	0.052386785	0.047184785	0.040925425	0.03095
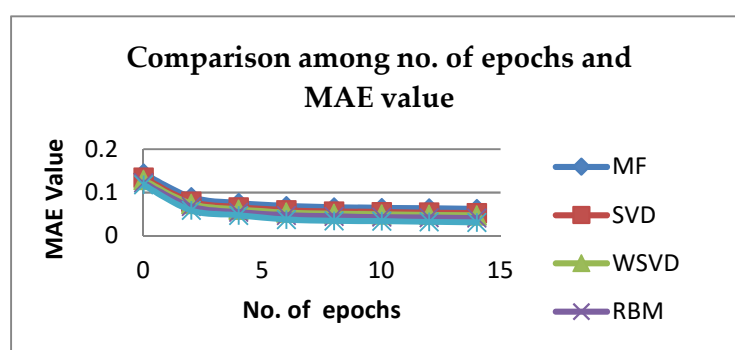
Figure 11. Graph between MAE and no. of epochs.

By analyzing and comparing all the approaches with proposed RBM-CNN method, this proposed method gives better accuracy considering two error measures such as RMSE and MAE. The integration of an RBM with CNN in a deep learning environment provides a better recommendation quality for choosing a hospital for a particular patient.

7. Conclusions and Future Work

Health Recommender systems are one of the new prevailing technologies for deriving supplementary information for a patient from healthcare data. These systems find recommended hospitals by calculating the similarity of patients' choices. Therefore, they play an important role in the medical sector. Modern state of the art technologies are required that can definitely resolve the issues data security and privacy found in CF-based health recommender systems. In this paper, different privacy-preserving collaborative filtering methods, along with the deep learning method, are compared. The proposed RBM-CNN demonstrates better accuracy of the health recommender system as compared to others. In the future, we will try to optimize our algorithm in order to provide better accuracy with a high level of privacy.

References

1. Isinkaye, F.O.; Folajimi, Y.O.; Ojokoh, B.A. Recommender systems: Principles, methods and evaluation. *Egypt. Inform. J.* **2015**, *16*, 261–273, ISSN 1110-8665. [CrossRef]
2. Burke, R.; Felfernig, A.; Goker, M.H. Recommender Systems: An Overview. *AI Mag.* **2011**, *32*, 13–18, ISSN 0738-4602. [CrossRef]
3. Li, S.; Kawale, J.; Fu, Y. Deep collaborative filtering via marginalized denoising auto-encoder. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2015; pp. 811–820.
4. Riyaz, P.A.; Varghese, S.M. A Scalable Product Recommenders using Collaborative Filtering in Hadoop for Bigdata. *Procedia Technol.* **2016**, *24*, 1393–1399. [CrossRef]
5. Priyadarshini, R.; Barik, R.K.; Panigrahi, C.; Dubey, H.; Mishra, B.K. An investigation into the efficacy of deep learning tools for big data analysis in health care. *Int. J. Grid High Perform. Comput.* **2018**, *10*, 1–13. [CrossRef]
6. Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Item-Based Collaborative Filtering Recommender Algorithms. In Proceedings of the ACM Digital Library International World Wide Web Conferences, Hong Kong, China, 1–5 May 2001; pp. 285–295, ISBN 1-58113-348-0.
7. Wang, Y.; Hajli, N. Exploring the path to big data analytics success in healthcare. *J. Bus. Res.* **2017**, *70*, 287–299. [CrossRef]
8. Babar, M.I.; Jehanzeb, M.; Ghazali, M.; Jawawi, D.N.A.; Sher, F.; Ghayyur, S.A.K. Big data survey in healthcare and a proposal for intelligent data diagnosis framework. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 7–12.
9. Lafta, R.; Zhang, J.; Tao, X.; Li, Y.; Tseng, V.S.; Luo, Y.; Chen, F. An intelligent recommender system based on the predictive analysis in the telehealthcare environment. *Web Intell.* **2016**, *14*, 325–336. [CrossRef]
10. Priyadarshini, R.; Barik, R.; Dubey, H. DeepFog: Fog Computing-Based Deep Neural Architecture for Prediction of Stress Types, Diabetes and Hypertension Attacks. *Computation* **2018**, *6*, 62. [CrossRef]
11. Martínez-Pérez, B.; De La Torre-Díez, I.; López-Coronado, M. Privacy and security in mobile health APPs: A review and recommendations. *J. Med. Syst.* **2015**, *39*, 181. [CrossRef]
12. Mu, R.; Zeng, X.; Han, L. A Survey of Recommender Systems Based on Deep Learning. *IEEE Access* **2018**, *6*, 69009–69022. [CrossRef]
13. Gope, J.; Jain, S.K. A survey on solving cold start problem in recommender systems. In Proceedings of the 2017 International Conference on Computing, Communication, and Automation (ICCCA), Greater Noida, India, 5–6 May 2017; pp. 133–138.
14. Jooa, J.H.; Bangb, S.W.; Parka, G.D. Implementation of a Recommender System usingAssociation Rules and Collaborative Filtering. *Inf. Technol. Quant. Manag. Procedia Comput. Sci.* **2016**, *91*, 944–952. [CrossRef]
15. Ponnam, L.T.; Punyasamudram, S.D. Health care Recommender System using Item Based Collaborative Filtering Technique. In Proceedings of the International Conference on Emerging Trends in Engineering, Technology and Science, Pudukkottai, India, 24–26 February 2016; Volume 1, pp. 56–60.

16.　Hoseini, E.; Hashemi, S.; Hamzeh, A. SPCF: A stepwise partitioning for collaborative filtering to alleviate sparsity problems. *J. Inf. Sci.* **2012**, *38*, 578–592. [CrossRef]

17.　Ma, X.; Lu, H.; Gan, Z.; Zeng, J. An explicit trust and distrust clustering based collaborative filtering recommender approach. *Electron. Commer. Res. Appl.* **2017**, *25*, 29–39. [CrossRef]

18.　Kaur, H.; Kumar, N.; Batra, S. An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system. *Future Gener. Comput. Syst.* **2018**, *86*, 297–307. [CrossRef]

19.　Archenaa1, J.; Mary Anita, E.A. Health Recommender System using Big data analytics. *J. Manag. Sci. Bus. Intell.* **2017**, *2*, 17–24.

20.　Behera, R.K.; Sahoo, A.K.; Pradhan, C.R. Big Data Analytics in Real Time—Technical Challenges and Its Solutions. In Proceedings of the 2017 International Conference on Information Technology (ICIT), Bhubaneswar, India, 21–23 December 2017; pp. 30–35. [CrossRef]

21.　Yang, C.C.; Jiang, L. Enriching User Experience in Online Health Communities Through Thread Recommendations and Heterogeneous Information Network Mining. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 1049–1060. [CrossRef]

22.　Cheng, H.T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anil, R. Wide & deep learning for recommender systems (DLRS 2016). In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016; pp. 7–10.

23.　Paul, P.K.; Dey, J.L. Data Science Vis-à-Vis efficient healthcare and medical systems: A techno-managerial perspective. In Proceedings of the 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 21–22 April 2017; pp. 1–8.

24.　Calero Valdez, A.; Ziefle, M.; Verbert, K.; Felfernig, A.; Holzinger, A. Recommender Systems for Health Informatics: State-of-the-Art and Future Perspective. In *Machine Learning for Health Informatics*; Holzinger, A., Ed.; Lecture Notes in Computer Science LNCS 9605; Springer: Cham, Switzerland, 2016.

25.　Sahoo, A.K.; Pradhan, C.R. A Novel Approach to Optimized Hybrid Item-based Collaborative Filtering Recommender Model using R. In Proceedings of the 2017 9th International Conference on Advanced Computing (ICoAC), Chennai, India, 14–16 December 2017; pp. 468–472.

26.　Baldominos, A.; De Rada, F.; Saez, Y. DataCare: Big Data Analytics Solution for Intelligent Healthcare Management. *Int. J. Interact. Multimed. Artif. Intell.* **2018**, *4*, 13–20. [CrossRef]

27.　Sharma, D.; Shadabi, F. The potential use of multi-agent and hybrid data mining approaches in social informatics for improving e-Health services. In Proceedings of the 4th IEEE International Conference on Big Data and Cloud Computing, Sydney, NSW, Australia, 3–5 December 2014; IEEE: New York, NY, USA, 2014; pp. 350–354.

28.　Harsh, K.; Ravi, S. Big Data Security and Privacy Issues in Healthcare. In Proceedings of the 2014 IEEE International Congress on Big Data, Anchorage, AK, USA, 27 June–2 July 2014; pp. 762–765.

29.　Portugal, I.; Alencar, P.; Cowan, D. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Syst. Appl.* **2018**, *97*, 205–227. [CrossRef]

30.　Li, T.; Gao, C.; Du, J. A NMF-based privacy-preserving recommender algorithm. In Proceedings of the 2009 First International Conference on Information Science and Engineering, Nanjing, China, 26–28 December 2009; pp. 754–757.

31.　Chen, J.; Li, K.; Rong, H.; Bilal, K.; Yang, N.; Li, K. A diagnosis and treatment recommender system based on big data mining and Cloud computing. *Inf. Sci.* **2018**, *435*, 124–149. [CrossRef]

32.　Fernández-Alemán, J.L.; Señor, I.C.; Lozoya, P.Á.O.; Toval, A. Security and privacy in electronic health records: A systematic literature review. *J. Biomed. Inf.* **2013**, *46*, 541–562. [CrossRef] [PubMed]

33.　Yuan, W.; Li, C.; Guan, D.; Han, G.; Khattak, A.M. Socialized healthcare service recommendation using deep learning. *Neural Comput. Appl.* **2018**, *30*, 2071–2082. [CrossRef]

34.　Wang, Y.; Kung, L.; Byrd, T.A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Chang.* **2018**, *126*, 3–13. [CrossRef]

35.　Zhou, X.; He, J.; Huang, G.; Zhang, Y. SVD-based incremental approaches for recommender systems. *J. Comput. Syst. Sci.* **2015**, *81*, 717–733. [CrossRef]

36.　Canny, J. Collaborative filtering with privacy via factor analysis. In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 11–15 August 2002; pp. 238–245.

37. Harper, F.M.; Konstan, J.A. The health care datasets: History and context. *ACM Trans. Interact. Intell. Syst.* **2015**, *5*, 19:1–19:19. [CrossRef]

38. Palanisamy, V.; Thirunavukarasu, R. Implications of big data analytics in developing healthcare frameworks—A review. *J. King Saud Univ.-Comput. Inf. Sci.* **2017**. [CrossRef]

39. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [CrossRef]

40. Ortega, F.; Hernando, A.; Bobadilla, J.; Kang, J.H. Recommending items to group of patients using Matrix Factorization based Collaborative Filtering. *Inf. Sci.* **2016**, *345*, 313–324. [CrossRef]

41. Ambika, M.; Latha, K. Intelligence Based Recommender System for Healthcare: A Patient-Centered Framework. In Proceedings of the 2nd International Conference on Advanced Theoretical Computer Applications, Ho Chi Minh City, Vietnam, 9–11 December 2015; pp. 245–255.

42. He, X.; Chua, T.S. Neural factorization machines for sparse predictive analytics. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 355–364.

43. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.S. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web (WWW '17), Perth, Australia, 3–17 April 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2017; pp. 173–182.

44. Wu, J.; Yang, L.; Li, Z. Variable Weighted BSVD-Based Privacy-Preserving Collaborative Filtering. In *Intelligent Systems and Knowledge Engineering (ISKE), Proceedings of the 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Taipei, Taiwan, 24–27 November 2015*; IEEE: Piscataway, NJ, USA, 2015; pp. 144–148.

45. Adams, R.J.; Sadasivam, R.S.; Balakrishnan, K.; Kinney, R.L.; Houston, T.K.; Marlin, B.M. PERSPeCT: Collaborative filtering for tailored health communications. In Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14), Foster City, Silicon Valley, CA, USA, 6–10 October 2010; pp. 329–332.

46. Sahoo, A.K.; Pradhan, C.; Mishra, B.S.P. SVD based Privacy Preserving Recommendation Model using Optimized Hybrid Item-based Collaborative Filtering. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 4–6 April 2019; pp. 294–298.

47. Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z. Collaborative filtering and deep learning Based recommender system for cold start items. *Expert Syst. Appl.* **2017**, *69*, 29–39. [CrossRef]

48. Dai, Y.; Wang, G. A deep inference learning framework for healthcare. *Pattern Recognit. Lett.* **2018**. [CrossRef]

49. Wang, X.; He, X.; Nie, L.; Chua, T.S. Item silk road: Recommending items from information domains to social users. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 185–194.

50. Jiang, L.; Yang, C.C. User recommendation in healthcare social media by assessing user similarity in heterogeneous network. *Artif. Intel. Med.* **2017**, *81*, 63–77. [CrossRef]

51. Yedder, H.B.; Zakia, U.; Ahmed, A.; Trajkovic, L. Modeling prediction in recommender systems using restricted boltzmann machine. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017.

52. Belle, A.; Thiagarajan, R.; Soroushmehr, S.M.; Navidi, F.; Beard, D.A.; Najarian, K. Big data analytics in healthcare. *Biomed Res. Int.* **2015**, *2015*, 370194. [CrossRef]