Due to the growth of IoT applications, especially health care, the information of patients’ health records using data collection from IoT-connected devices has been considered. Biological data of patients in the health record helps to monitor the patient’s status and identify various diseases. Chronic diseases are a type of silent disease that, if not diagnosed in time, can cause irreparable damage to patients. The use of patients’ medical record data for early diagnosis of chronic diseases has recently attracted the attention of many researchers. On the other hand, the application of machine learning methods in the form of recommender systems has taken an important step in improving medical services and health care. In this paper, a medical recommender system was presented to identify and treat chronic diseases using an IoT device. In the present method, the electronic patient health record dataset that is loaded in the PhysioNet data repository has been used. In the present dataset, patients’ health records have been recorded according to the identified diseases and the physician’s diagnosis. In the proposed method, the $K$-nearest neighbor classification method is used to identify the type of disease, and the collaborative filtering method is used to find the appropriate treatment for patients. The results of the implementation of the proposed method show that this approach, based on the use of symptom similarity among patients, has good accuracy in diagnosing and predicting chronic diseases and has provided higher results than previous methods.

1. Introduction

With the abundant volume of IoT-related datasets, extracting useful information from this data is a challenging task. Especially in the healthcare sector, there is a huge amount of data related to patients’ tests, doctors’ opinions, types of diagnosed diseases, and medications needed by patients. Therefore, finding data on the treatment of a particular disease and specialists’ opinions about it has become a hard issue. Therefore, having guidelines for using this data can be useful. In IoT healthcare systems, recommenders play a guiding role. Recommendation systems have been developed to provide suggestions to users and companies or organizations to eliminate the challenge of data volume and the existence of similar information. Recommenders are adjusted based on the user’s preferences and try to provide
useful information according to the experience of other users as well as the user’s goals [5–9].

The Internet of Things has ushered in a new era in the healthcare industry by integrating medical equipment. Patient data can be collected based on sensors embedded in IoT-connected equipment. However, data manipulation and the effective use of patient information require a precise mechanism. In health care, effective diagnosis of the disease based on patient data and finding appropriate treatment remains an important challenge. Recommender systems are a critical step in promoting IoT benefits. Recommender systems generally include procedures that identify the user’s needs based on the information available and provide the best solution for the user. Given the vast amount of information available through the Internet of Things, patients are likely to receive a variety of recommendations for services or treatment of the disease. Therefore, choosing between these approaches requires a precise mechanism and effective recommending system [10–13].

Over the past few years, we have seen some remarkable studies on recommendation systems. In [14], an IoT-based efficient community-based recommender system for diagnosing the types of cardiac disease and suggested recommendations related to the treatment has been proposed. In [15], an ontology-driven personalized food recommendation in IoT-based healthcare system has been proposed. In [16], a recommender system using software-defined networking in IoT-based smart healthcare ecosystem has been designed. In [17], a recommender system was introduced which was intended for patient’s behavior correction based on gathering various health state parameters and lifestyle-related data. In [18], an integrated recommendation system into connected health has a design that will expand on the opportunities for better data accessibility and use them for chronic diseases. In [19], a framework has been developed in which decision tree and support vector machine (SVM) techniques were used for predicting hypertension.

Recommended systems have been increasingly used by many applications such as e-commerce and e-health due to their ability to automatically extract useful information and predict and recommend appropriate results to consumers [7, 8, 20]. Therefore, a referral system for chronic disease management is needed to minimize both the risk of such a disease and its cost [21–24].

In recent years, a pattern of chronic diseases in the Middle East, like the rest of the world, has begun to emerge [25, 26]. These diseases appear in the form of obesity, heart disease, diabetes, and other chronic diseases. This increase in chronic illnesses, coupled with an increase in inactive lifestyles in the Middle East, has put a lot of pressure on healthcare providers, especially when trying to ensure patient follow-up that occurs after each treatment change [27]. The rapid development of information and communication technology has created a new era for researchers to play a number of e-health programs that play an important role in improving health services [28].

Therefore, in this study, patients’ health records have been used in order to create a referral system for diagnosing and predicting chronic diseases. In the present study, the electronic patient health record dataset uploaded to the PhysioNet data repository will be used. In the present dataset, the patients’ health records have been recorded according to the identified diseases and the physician’s diagnosis. In the proposed method, symptoms related to various diseases are trained based on the type of disease extracted from the dataset by the nearest neighbor classification method. In the nearest neighbor classification method, patients are classified according to the symptoms of the disease, and the registered diagnosis for these patients is presented as a recommendation in the proposed medical advisor system. In order to evaluate the proposed method, the criteria of accuracy, sensitivity, and accuracy, the recommendations provided will be used, and eventually, the proposed method will be compared with other existing methods in this field. It is expected that the proposed method, based on the use of similarity of symptoms among patients, will have good accuracy in diagnosing and predicting chronic diseases.

The purpose of this paper in the first step is to find the type of disease based on the symptoms of patients collected through their medical records. Then, the next step is to identify the stage of disease progression and the necessary treatment for it according to the medical records of other patients in this field as well as the opinions of experts. In fact, the recommending system in the proposed method first tries to diagnose the type of chronic disease and then, by finding similar patients with similar symptoms from the archives of medical records, uses the opinions of their treating physician and its specific drugs.

The motivation of this article is that it is difficult to diagnose chronic diseases due to the fact that chronic diseases grow secretly, and most patients do not have any specific symptoms. On the other hand, the treatment of chronic diseases requires a lot of expertise and experience from the doctor. Patients with chronic diseases may be at risk based on misdiagnosis. Therefore, the existence of a platform to share the experience of physicians specializing in chronic diseases seems necessary. In this paper, this platform is introduced as a medical referral system. This recommender system consists of two stages. In the first stage, according to the patients’ health records, the type of chronic disease is diagnosed. Given that laboratory results vary in chronic diseases, the best way to diagnose the type of disease is to compare the patient’s laboratory data with the types of patients with chronic diseases. So we have a multiclass problem that every class is a kind of chronic disease. In the present article, the nearest neighbor method has been used to classify new patients. In this method, the characteristics of each patient’s medical record are compared with other patients in the dataset, and based on the similarity to patient K, the new patient class is determined. After determining the type of chronic disease, in the next step, we will determine the treatment required for chronic disease. At this stage, according to the treatment prescribed for similar patients based on the participatory filter method, we will recommend treatment for new patients to treating physicians. According to this method, each patient receives several treatment recommendations according to the previous similar patients, which will
determine the majority vote in the appropriate treatment for the patient.

The main contribution of the article is summarized as follows:

(i) Diagnosis of the type of chronic disease using the nearest neighbor classification method based on data collected from medical devices in the Internet of Things that are stored in patients’ health records

(ii) Examine the medical records of all patients for accurate diagnosis of the disease

(iii) Identify patients similar to the new patient in a specific type of chronic disease based on the patient’s symptoms and physicians’ supervision to determine the stage of disease progression

(iv) Recommendations for early treatment of chronic disease based on the health records of similar patients using the collaborative filtering method

In the rest of this paper, in the second part, clustering methods based on collective intelligence in the wireless sensor network will be examined. In the third section, the proposed method will be described in detail. In the fourth section, the test results and evaluation of the proposed method will be presented. In the fifth section, conclusions and future work will be presented.

2. Related Works

Due to the importance of early diagnosis of chronic diseases, many researchers have tried to provide medical recommendation systems, whose studies will be briefly reviewed.

In 2020, Mustaqeem et al. developed a common filtering method based on multistage clustering. This method is applied in the monitored dataset for four different types of cardiovascular disease. Patient data are categorized according to their respective disease classes based on k-means clustering. A query patient who is once directed to the correct division of the disease needs to obtain similarity scores from a reduced subcluster, thus improving system performance. Each disease partition has a separate recommendation process that modulates and helps improve system scalability [29].

In 2020, Saha et al. provide an overview of the challenges associated with existing medical referral systems. By improving machine learning techniques, the referral system offers various opportunities for medical science. Systems can operate more efficiently through deep learning and solve complex problems, even when the dataset is diverse and unstructured [30].

In 2018, Chen et al. introduced the DDTRS system to maximize the use of advanced medical technology in advanced hospitals and the rich medical knowledge of experienced physicians. First, to accurately identify the symptoms of the disease, the density-peaked clustering analysis (DPCA) algorithm for clustering the symptoms is introduced. In addition, communication analysis of disease diagnosis rules and disease treatment rules was performed separately by the Apriori algorithm [23].

In 2018, Jabeen et al. developed an efficient IoT-based referral system that diagnoses heart disease and its type and provides physical and dietary advice. The first part intends to collect patient data remotely using biosensors. The cardiovascular prediction model is then implemented, which can diagnose cardiovascular disease and classify it into eight existing cardiovascular classes. The second part is to provide physical and diet plan recommendations to the heart patient according to gender and age groups [14].

In 2015, Kamran et al. presented the classification of medical recommendation systems among content-based, participatory, demographic, knowledge-based, and hybrid techniques. The proposed work focuses on providing an overview of referral systems in health care [31].

In 2012, Hussein et al. proposed a recommending system approach based on a hybrid method using multiple classifications and integrated shared filtration. The purpose of this article is to provide more accurate and effective recommendations to help patients control their chronic illness and help healthcare providers to have a 24-hour remote patient monitoring system [1]. In 2014, they also introduced a chronic disease referral system that helps patients control their disease through medical advice and diagnosis. For system accuracy, which is critical in such an application, high-dimensional data must be handled [32].

In [33], effective mechanism for the protection of digital library readers’ lending privacy under a cloud environment has been presented. In [34], a large number of research achievements relevant to user privacy protection in an untrusted network environment has been reviewed. In [35], a group of dummy query sequences, to cover up the query locations and query attributes of mobile users and thus protect users’ privacy in LBS has been constructed. In [36], a time-aware deep CF framework which contains dynamic user preference modeling based on attention mechanism and matching score prediction based on DL has been presented. In [37], a location privacy-preserving system for LBS by constructing “cover-up ranges” to protect the query ranges associated with a location has been proposed. In [38], a group of plausible fake queries for each user book query to cover up the sensitive subjects behind users’ queries has been constructed. In [38], a new shapelet discovery method, referred to as Pruning Shapelets with Key Points (PSKP), has been proposed. In [39], a group of fake preference profiles, to cover up the user-sensitive subjects and thus protect user personal privacy in personalized recommendation, has been generated. In [40], this approach to protect the preference privacy behind users’ book browsing behaviors in a digital library has been designed. In [41], a new classification approach based on Wikipedia matching has been proposed. In [42], a topic modeling-based approach to extractive automatic summarization has been presented. In [43], a client-based approach to address plausible but innocuous pseudoqueries together with a user query has been proposed. In [44], a new contextual advertising approach has been proposed, which uses Wikipedia thesaurus knowledge to enrich the semantic expression of a target page.
3. Methodology

In this section, we will first review the prerequisites of the proposed method. The details of the proposed method will then be described.

3.1. Preliminaries of the Proposed Method

3.1.1. K-Nearest Neighbor (KNN). The K-nearest neighbor classification (KNN) method is a group of K records in the training suite that are close to the tested data and, based on the label assignment, find the specific class that dominates the neighborhood. This method demonstrates the fact that in many datasets, one record is unlikely to correspond exactly to another, as well as the fact that conflicting information about the class of a record can result from records close to it must be presented [45, 46]. There are several key elements to this approach:

(1) A set of tagged records is used to evaluate the record class being tested

(2) To calculate the proximity of records, the criterion of distance or similarity can be used

(3) K value, the number of nearest neighbors, must be specified

(4) The method used to determine the target record class should be determined based on the classes and the distance K of the nearest neighbor

In its simplest form, KNN can assign a class record to its nearest neighbors or most of its nearest neighbors [46].

Several key issues affect KNN performance, one of which is the choice of k. Figures 1–3 show an exam record with an x tag and training records belonging to each of the “-“ or “+“ classes. If k is too small, the result can be sensitive to noise points. On the other hand, if k is too large, the neighborhood may contain many points from many other classes. The estimation of the best value for k can be obtained by cross-validation. In this method, the value of K starts from a small value, and a constant value is added to it for each iteration. The performance of the algorithm is compared in each iteration, and finally, the value of k, which leads to the best performance for the algorithm, is chosen as the principal value for k.

However, it is important to note that k = 1 may also have other values of k, especially for small datasets commonly used in research or classroom exercises. However, with sufficient records, values greater than k are more resistant to noise.

Another issue is the approach of combining class labels. The easiest way is to pay attention to the majority vote, but this can be problematic if the nearest neighbors are very diverse in proximity, and the nearest neighbors are more confident of the target class. A more sophisticated method, which is usually less sensitive to the selection of k, is to measure each record based on its distance. Various options are possible. For example, the weight coefficient is taken from the square of the mutual distance \( w_j = 1/d(y, z)^2 \). This value replaces the last step of the basic KNN algorithm as follows [46]:

\[
\text{Distance – Weighted Voting} : c_z = \arg\max_v \sum_{j \in N} w_j \times I(v = \text{class}(c_j)).
\] (1)

Selecting the distance criterion is another important point to consider in this method. The Euclidean distance or Manhattan distance is usually used as the distance criterion in this method. For two points, x and y, with n properties, this distance is calculated by the following formulas:[46]:

\[
d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} \text{Euclidan distance},
\] (2)

\[
d(x, y) = \sum_{k=1}^{n} |x_k - y_k| \text{Manhattan distance},
\]

where \( x_k \) and \( y_k \) are the attributes k, x, and y, respectively.

Although these and various other criteria can be used to calculate the distance between two points, conceptually, the most desirable measure is the distance that is less or more similar between two records, meaning that using this, the criteria probably have the same class tags between the two records, so, for example, if KNN is used to classify documents, it is better to use the cosine criterion instead of the Euclidean distance. Note that KNN can also use data and categories for data with classified or mixed properties until an appropriate distance criterion is defined.

3.1.2. Collaborative Filter. Collaborative filtering (CF) is one of the technologies that has been used in many cases in the recommendation. This method calculates the similarity between users and uses this information to recommend products that have not yet been reviewed by the user. The similarity is based on past reviews of shared items. This similarity is used to generate recommendations for items that have already been reviewed by similar users but have not yet been specified by the user [47, 48].

The first step of the CF system is to search for users with similar consumption habits, for example, to calculate the similarity between users. When analyzing users 1 and 5, for option I₁, the difference between their rankings is 1.0. In I₂, there is no difference between them, and for I₃, the difference is 0.5.

If the difference between the two users’ ratings on the options is less than a threshold value, it can be said that the two users are similar. Hence, it can be said that users 1 and 5 are similar. By the same token, users 1 and 2 will not be the same. The similarity calculation can only occur for the options that both users have preferably stated. There are several techniques for calculating similarities
between users. The most common methods are shown in Equation (3) [17]:

\[ \hat{p}_{a,u} = \bar{r}_a + \frac{\sum_{i=1}^{h} (r_{a,i} - \bar{r}_a) \cdot w_{a,u}}{\sum_{i=1}^{h} |w_{a,u}|}. \] (3)

(i) \( w_{a,u} \) is the relationship of the target user to a given user \( U \)

(ii) \( r_{a,i} \) is the target user score for an \( i \) option

(iii) \( \bar{r}_a \): the average of all rankings is a target user \( a \)

(iv) \( \hat{p}_{a,u} \) is the expected use of the option for user \( u \)
number of samples and then combining their results. The data mining team can significantly improve the prediction of people at risk for chronic disease. Therefore, predicting people with this type of kidney disease is an important issue that should be considered.

3.2. Proposed Method. As mentioned, efforts to predict disease in terms of providing medical services and making the best possible decision on disease samples are very important for hospitals and medical institutions. Failure to make the right decision and take timely action to treat some diseases can pose an irreparable risk to the patient or in some cases the patient may die due to lack of medical care. Chronic kidney disease is one of the diseases that, if not detected and treated in the early stages, can have irreversible consequences for the patient’s health, even if forgotten can lead to death. Therefore, predicting people with this type of kidney disease is an important issue that should be considered.

Therefore, in order to solve this problem and predict patients with chronic diseases, a number of predictive models have been proposed. Recently, the data mining group has been used successfully in a variety of applications, including to assist in medical diagnosis and disease prediction. The data mining team can significantly improve the ability to generalize learning systems by training a limited number of samples and then combining their results.

Therefore, in this study, we try to use the nearest neighbor to use standard data from the biological test results of some patients with chronic disease, to predict people at risk of chronic disease. The use of the nearest neighbor in disease prediction has been proven in many publications, so in this study, the use of this classification has been confirmed. Figure 1 shows the overall architecture of the proposed method.

In this study, we want to predict people with these diseases using the nearest neighbor model based on the medical consulting system. In this regard, the data obtained from the test results, which is a standard database in the standard PhysioNet data repository, are entered as artificial data into artificial neural networks and processed by zinc. These data predict the outcome of people at risk for chronic disease.

Due to the different amounts of data and the variety of features in the experimental results used in this study as the nearest neighbor input, the number of features in the dataset or in other words the dimensions of the data is high. The large size of the data imposes more complexity on the system and may affect the performance accuracy of the classification model. Therefore, in this research, we first select the feature subset from the preprocessing step to reduce the data size. It is used to reduce the volume of calculations. After this preprocessing step, features that are less relevant to the class tag are removed, and the remaining features are important features that play a key role in teaching the model.

Another step of preprocessing used in this research is to convert nonnumerical data to numerical data. Because the nearest neighbors perform better with numerical data, this preprocessing step is necessary to prepare the data for model training. In this preprocessing step, nonnumeric data are replaced with numbers that represent the original data using the variable change method and increase the ease of model training to classify the nearest neighbor. The output data from these preprocessing stages are model-ready data that is injected as input data to the nearest neighbor. The flow diagram of the proposed method is shown in Figure 2.

As shown in Figure 2, the proposed method includes the following steps:

Data extraction: in this step, information about the status of patients with chronic diseases is extracted from patients’ health records from the proposed method. This information includes biological tests of patients that were collected during the treatment process in the hospital and recorded in the patients’ health records. There are also ways to treat diseases in patients’ health records. Thus, patients with chronic diseases are used as input to the proposed referral system according to their symptoms.

Data classification: due to the fact that the nearest neighbor classifier has been identified as a lazy and sample-based classification, the training process is delayed until the test data is entered. Therefore, in the proposed method, the training process will take place at the time of the new patient’s visit to the hospital. In the process of teaching the biological signs, characteristics of new patients are compared with all patients in the dataset. In comparison between properties, since the values of the properties have been converted...
In this paper, the max-min normalization method is used to normalize the values of patient characteristics. This method is used to divide the interval and the direct use of this data may a provide medical recommendation, similar patients should be identified from all patients with the same type of disease. Given that in chronic patients the symptoms of different symptoms as the disease progresses, the best way to find the type of disease is to compare a new sample with all N previous patients. Given that the KNN classifier inherently compares a test sample with all training samples to find the most similar ones, this classifier is the best option for our work. It is assumed that we have L new patients and each of the patients has M feature that is stored in their electronic health record.

Given that the values of features may be in different intervals and the direct use of this data may affect the calculation of similarity between new patients and old patients, the values of patient characteristics should be normalized. In this paper, the max-min normalization method is used [49], which is shown in

$$\text{NormalizedData} = \frac{X(i) - \min (X)}{\max (X) - \min (X)}.$$  \hspace{1cm} (4)$$

$X(i)$ represents the original data value, $\min (x)$ the lowest possible value for the data, and $\max (X)$ the highest possible value for the data. So, the similarity of a new patient to other patients is obtained based on Equation (5).

$$\text{Sim}_{n,j} = 1 - \sum_{i=1}^{L} \sum_{j=1}^{N} \sum_{f=1}^{M} \sqrt{(x_{i,j} - x_{f,j})^2}, \hspace{1cm} (5)$$

where $x$ represents the value of the feature, $f$ represents the feature index in the dataset, $i$ represents the index of new patients, and $j$ represents the index of previous patients. Based on Equation (4), $N$ values have obtained for each new patient, which indicates the similarity with the $N$ of the previous patient. Among these $N$ patients, $K$ patients with the most similarity to the new patient are selected as nearest neighbors. In this article, the value of $k$ is considered equal to 7. Finally, the type of disease associated with a new patient is determined by voting between $K$ nearest neighbors.

**Providing medical recommendation:** at this stage of the proposed method, according to the type of disease determined for the new patient and according to the health records of neighbors close to the new patient, treatment advice is provided to the new patient. After diagnosing the type of chronic disease in a new patient, in order to provide medical recommendation, similar patients should be identified from all patients with the same type of chronic disease. Patients with a specific type of chronic disease may have different symptoms as the disease progresses. In addition, each symptom may have a different effect on the course of the disease. Therefore, physicians’ opinions on patients’ characteristics are added to the patients’ health records as a weight. Therefore, in order

### Table 1: Calculate the distance of a part of the test patients from the educational patients.

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 6</th>
<th>Test 7</th>
<th>Test 8</th>
<th>Test 9</th>
<th>Test 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4055</td>
<td>0.4274</td>
<td>0.3620</td>
<td>1.9725</td>
<td>0.3592</td>
<td>0.3569</td>
<td>0.8528</td>
<td>1.3559</td>
<td>1.3559</td>
<td>0.7225</td>
</tr>
<tr>
<td>0.2880</td>
<td>0.3224</td>
<td>0.2445</td>
<td>1.9860</td>
<td>0.3190</td>
<td>0.3185</td>
<td>0.8094</td>
<td>1.3334</td>
<td>1.3334</td>
<td>0.8909</td>
</tr>
<tr>
<td>0.5994</td>
<td>0.6148</td>
<td>0.5566</td>
<td>2.0425</td>
<td>0.5027</td>
<td>0.5000</td>
<td>0.9786</td>
<td>1.4425</td>
<td>1.4425</td>
<td>0.5130</td>
</tr>
<tr>
<td>1.0276</td>
<td>1.0371</td>
<td>0.9821</td>
<td>2.2129</td>
<td>0.8942</td>
<td>0.8916</td>
<td>1.2847</td>
<td>1.6658</td>
<td>1.6658</td>
<td>0.2520</td>
</tr>
<tr>
<td>0.5511</td>
<td>0.5695</td>
<td>0.5025</td>
<td>2.0425</td>
<td>0.4583</td>
<td>0.4560</td>
<td>0.9457</td>
<td>1.4213</td>
<td>1.4213</td>
<td>0.5647</td>
</tr>
<tr>
<td>0.2899</td>
<td>0.3248</td>
<td>0.2441</td>
<td>1.9998</td>
<td>0.3189</td>
<td>0.3185</td>
<td>0.8289</td>
<td>1.3483</td>
<td>1.3483</td>
<td>0.8907</td>
</tr>
<tr>
<td>0.3150</td>
<td>0.3443</td>
<td>0.2766</td>
<td>1.9860</td>
<td>0.3209</td>
<td>0.3196</td>
<td>0.8428</td>
<td>1.3554</td>
<td>1.3554</td>
<td>0.8308</td>
</tr>
<tr>
<td>1.6000</td>
<td>1.5697</td>
<td>1.7251</td>
<td>0.9812</td>
<td>1.7420</td>
<td>1.7376</td>
<td>1.4257</td>
<td>1.6858</td>
<td>1.6858</td>
<td>2.0084</td>
</tr>
<tr>
<td>1.5957</td>
<td>1.5654</td>
<td>1.7206</td>
<td>0.9812</td>
<td>1.7376</td>
<td>1.7332</td>
<td>1.4202</td>
<td>1.6814</td>
<td>1.6814</td>
<td>2.0045</td>
</tr>
<tr>
<td>1.6000</td>
<td>1.5697</td>
<td>1.7251</td>
<td>0.9812</td>
<td>1.7420</td>
<td>1.7376</td>
<td>1.4257</td>
<td>1.6858</td>
<td>1.6858</td>
<td>2.0084</td>
</tr>
<tr>
<td>1.6000</td>
<td>1.5697</td>
<td>1.7251</td>
<td>0.9812</td>
<td>1.7420</td>
<td>1.7376</td>
<td>1.4257</td>
<td>1.6858</td>
<td>1.6858</td>
<td>2.0084</td>
</tr>
<tr>
<td>1.3823</td>
<td>1.3900</td>
<td>1.1813</td>
<td>1.9614</td>
<td>1.0255</td>
<td>1.0244</td>
<td>1.5910</td>
<td>1.9621</td>
<td>1.9621</td>
<td>0.7575</td>
</tr>
<tr>
<td>1.3820</td>
<td>1.3896</td>
<td>1.1813</td>
<td>1.9594</td>
<td>1.0254</td>
<td>1.0243</td>
<td>1.5910</td>
<td>1.9620</td>
<td>1.9620</td>
<td>0.7576</td>
</tr>
<tr>
<td>1.3817</td>
<td>1.3892</td>
<td>1.1813</td>
<td>1.9574</td>
<td>1.0254</td>
<td>1.0242</td>
<td>1.5910</td>
<td>1.9620</td>
<td>1.9620</td>
<td>0.7576</td>
</tr>
<tr>
<td>1.3821</td>
<td>1.3896</td>
<td>1.1814</td>
<td>1.9595</td>
<td>1.0255</td>
<td>1.0244</td>
<td>1.5911</td>
<td>1.9621</td>
<td>1.9621</td>
<td>0.7575</td>
</tr>
<tr>
<td>1.4330</td>
<td>1.4403</td>
<td>1.3967</td>
<td>1.9931</td>
<td>1.4243</td>
<td>1.4244</td>
<td>1.0161</td>
<td>0.0690</td>
<td>0.0690</td>
<td>1.7213</td>
</tr>
<tr>
<td>1.4351</td>
<td>1.4415</td>
<td>1.3991</td>
<td>1.9545</td>
<td>1.4264</td>
<td>1.4264</td>
<td>1.0204</td>
<td>0.0142</td>
<td>0.0142</td>
<td>1.7238</td>
</tr>
<tr>
<td>1.4331</td>
<td>1.4402</td>
<td>1.3968</td>
<td>1.9844</td>
<td>1.4244</td>
<td>1.4244</td>
<td>1.0165</td>
<td>0.0515</td>
<td>0.0515</td>
<td>1.7215</td>
</tr>
<tr>
<td>1.4349</td>
<td>1.4413</td>
<td>1.3989</td>
<td>1.9563</td>
<td>1.4262</td>
<td>1.4262</td>
<td>1.0200</td>
<td>0.0111</td>
<td>0.0111</td>
<td>1.7236</td>
</tr>
<tr>
<td>1.4344</td>
<td>1.4410</td>
<td>1.3984</td>
<td>1.9609</td>
<td>1.4258</td>
<td>1.4257</td>
<td>1.0192</td>
<td>0.0079</td>
<td>0.0079</td>
<td>1.7231</td>
</tr>
</tbody>
</table>
to find patients similar to the new patient, Equation (6) has been used.

\[
\text{sim}_{ij} = \sum_{r=1}^{M} \left( r_{j} + \frac{\sum_{h=1}^{h} (r_{ij} - \bar{r}_{j}) \ast w_{ij}}{\sum_{h=1}^{h} |w_{ij}| + |w_{ij}|} \right),
\]

(6)

where \( r_{j} \) represents value of feature for the new patient, \( h \) is the number of patients with a specific type of chronic disease, \( r_{ij} \) is the difference in the feature value for the new patient and the previous patients, \( \bar{r}_{j} \) is the average feature value for the whole patient, \( w_{ij} \) is the difference in weight of feature for the new patient and the previous patient, and \( w_{ij} \) is the weight of feature for new patient.

Finally, \( \text{sim}_{ij} \) which is the amount of similarity of the new patient to each of the previous patients is obtained based on the collaborative filter and the opinions of experts. Patients have the most similarity and are selected as target patients. Medical recommendation is provided to new patients based on the treatment method and medical instructions included in the medical bird of the target patients.

Chronic patients are one of the most common patients, in which progress without pain in the body to eventually cause great harm to the patient. Chronic disease control requires accurate and early diagnosis. On the other hand, due to the different symptoms in different chronic diseases and even in a specific type of disease, in the diagnosis of chronic diseases, it is necessary to examine the symptoms related to different diseases. One of the significance of the proposed method is to use the KNN classifier to diagnose the type of disease. Due to its instance-based nature, this classification examines the symptoms of all diseases to obtain the most similar type of disease for a new patient. On the other hand, in chronic disease of a certain type, different symptoms in patients may be related to the stage of disease progression, while some novice physicians may diagnose them as different diseases. Therefore, it is necessary to examine similar patients in different stages of the patient. Another significance of the proposed method is to use a collaborative filter to find similar patients based on the patient’s characteristics and symptoms and the opinion of physicians, among all patients with a specific type of chronic disease.

4. Implementing the Proposed Method

To implement the K-NN classification model, we must first normalize the data from the training and test datasets and then calculate the distance of the tested patients in terms of similarity criteria to the training patients. As mentioned earlier, the Euclidean distance criterion is used to calculate the similarity interval between patient education and testing in the K-NN classification model. This distance criterion will be applied in this study to the features that were identified in the previous section as similarity criteria between patients. The similarity criterion is actually the characteristics that are directly related to the class label and can determine the group of each patient individually or in combination with some other characteristics. As mentioned earlier, the Euclidean distance criterion deals with numerical data, and the required similarity criteria must be in the form of numerical data in order to obtain accurate results. In this study, all values related to the properties have numerical values.

By determining the distance of experimental patients from educational patients, \( K \) educational patients who are less distant from the experimental patient in terms of similarity criteria, in other words, educational patients who are more similar to the experimental patient, are selected as close neighbors. The value of \( k \) is a criterion for selecting the number of neighbors that must be voted on to determine the patient class, which is selected by default for the user. If \( k \) has a small value, for example 1 or 2, the test patient class may not be determined correctly, and if \( k \) has a large value such as 9 or 10, the test patient class may be affected by noise and unrelated information in the neighborhood. Be impressed. Therefore, \( k \) should have an optimal value between 3 and 7, and in this study, the value of \( k \) should be considered equal to 5.

By selecting \( k \) as the nearest neighbor of the tested patients, the class label of this neighbor \( k \) is checked. Any class that has more representation among these close neighbors is considered the dominant class, and the patient class label is easily determined in the test. Table 1 shows the process of calculating the distance of experimental patients from patient education for 200 experimental patients.

As shown in Table 1, the distance between the tested patients and the trained patients was calculated in terms of similarity criteria for 10 experimental patients compared to 20 trained patients. The columns in Table 1 include the tested patients, and the rows include the patients under training. The last column of Table 1 contains the label of the educational class of educational patients, which is determined by determining the distance, the label of the patient’s medical class, which is considered a close neighbor.

According to Table 1, it can be seen that the educational patients are italicized with the shortest distance compared to the tested patients. In each column, 5 patients are identified, representing 5 neighbors close to the experimental patient in that column. The class label of these tested patients is determined by determining the class label of the nearest neighboring educational patients and voting among these educational patients. In such a way that the representative of each class that is more among these 5 neighbors is recognized as the dominant class, and that class is selected as the experimental patient class. In some patients, it is observed that all the neighbors are from the same class where the class label is easily marked, but in some other patients, other members of the class are seen among the neighbors. In this case, determining the patient’s class label is a bit difficult. Table 2 shows the trend of assigning experimental patients to chronic diseases in close neighbors.

As shown in Table 2, the patient distance applies to all close neighbors of the tested patients, and the closer distance for the patient test has a higher membership rate. This degree of membership is in fact the probability of an experimental patient belonging to a chronic disease according to the educational patient under study. The higher the degree
of membership of an educational patient, it means that the experimental patient, due to the similarity of the symptoms with the educational patient, most likely (equal to the degree of membership of the educational patient) belongs to a chronic disease.

4.1. Provide Medical Recommendations. According to the results of the nearest neighbor classification to provide medical advice to the Bibmaren test according to the health records of patients, the medical advice provided in the proposed method includes the same patient, according to which the new patient can be treated according to the patient’s medical record. The difference between the predicted value and the actual rating value is recognized by the user patient as a proven error, which is used as one of the evaluation criteria of the proposed systems. Table 3 shows a sample of results for similar patients with respect to similarity options and the type of disease diagnosed for that patient, as well as the amount of error associated with predicting one patient for other educational patients similarly shown.

As shown in Table 3, similar patients were calculated according to the common characteristics for each patient extracted from the educational dataset and the error related to each user. In the continuation of this chapter, we will evaluate the proposed method.

4.2. Evaluate the Proposed Method. The performance quality of a medical consulting system can be measured by many criteria. The type of criteria used depends on the method used. The performance of the medical counseling system is determined by the accuracy of the recommendations provided to patients. In this study, considering the similar method of close neighbors, two criteria have been used to measure the performance accuracy of the proposed method. The first criterion is statistical accuracy to measure the accuracy of the predicted disease for patients, and the second criterion is the accuracy in providing counseling to patients. Finally, the performance accuracy of the proposed method is compared with previous methods.

4.3. Performance Evaluation of the Proposed Method Based on Statistical Accuracy. As mentioned, statistical accuracy measures examine the direct comparison of patients’ predictions for patients with their actual disease type. Mean absolute error (MAE) and root mean square error (RMSE) are commonly used as the measures of statistical accuracy. MAE is one of the most common methods used among most

<table>
<thead>
<tr>
<th>Test</th>
<th>K</th>
<th>Distance</th>
<th>Prob</th>
<th>Predicted class</th>
<th>Test</th>
<th>K</th>
<th>Distance</th>
<th>Prob</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.2998</td>
<td>0.8477</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0.318</td>
<td>0.8425</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.3002</td>
<td>0.8473</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0.3191</td>
<td>0.8423</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.3033</td>
<td>0.8442</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>0.3210</td>
<td>0.8405</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.3446</td>
<td>0.7988</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>0.3593</td>
<td>0.8001</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.4465</td>
<td>0.6622</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>0.4584</td>
<td>0.6746</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.2521</td>
<td>0.9635</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>0.9812</td>
<td>0.8745</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.5131</td>
<td>0.8489</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>0.9812</td>
<td>0.8745</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.5647</td>
<td>0.8169</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>0.9812</td>
<td>0.8745</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.7225</td>
<td>0.7003</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>0.9812</td>
<td>0.8745</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0.7576</td>
<td>0.6705</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>1.9546</td>
<td>0.5020</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.08</td>
<td>0.9919</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0.2442</td>
<td>0.8971</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.0112</td>
<td>0.9841</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>0.2446</td>
<td>0.8968</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.0142</td>
<td>0.9740</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>0.2766</td>
<td>0.8680</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.0515</td>
<td>0.6603</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>0.3621</td>
<td>0.7738</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.069</td>
<td>0.3897</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>0.5026</td>
<td>0.5643</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.8095</td>
<td>0.8217</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>0.3224</td>
<td>0.8755</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.8290</td>
<td>0.8130</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>0.3248</td>
<td>0.8737</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.8428</td>
<td>0.8067</td>
<td>0</td>
<td>9</td>
<td>3</td>
<td>0.3444</td>
<td>0.8580</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.8528</td>
<td>0.8021</td>
<td>0</td>
<td>9</td>
<td>4</td>
<td>0.4275</td>
<td>0.7812</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.9457</td>
<td>0.7566</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>0.5695</td>
<td>0.6116</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.3185</td>
<td>0.8416</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>0.2880</td>
<td>0.8870</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.3186</td>
<td>0.8415</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>0.2899</td>
<td>0.8856</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0.3196</td>
<td>0.8405</td>
<td>0</td>
<td>10</td>
<td>3</td>
<td>0.3151</td>
<td>0.8649</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.3569</td>
<td>0.8011</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>0.4055</td>
<td>0.7761</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.4561</td>
<td>0.6752</td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0.5512</td>
<td>0.5864</td>
<td>1</td>
</tr>
</tbody>
</table>
methods. Naturally, the lower the value of this criterion for each patient, the higher the accuracy of the recommended method. The relationship of this criterion is calculated as follows:

\[
MAE = \frac{1}{N} \sum_{u,j} \left| p_{u,j} - r_{u,j} \right|,
\]

where \( p_{u,j} \) is defined as the predicted disease for patient \( u \) based on the symptoms \( i \), \( r_{u,j} \) is the actual disease of the patient \( u \) based on the symptoms \( i \), and \( N \) is the total number of patients in the dataset. Figure 3 shows the average error rate for 50 patients.

As shown in Figure 3, the average error in ranking prediction for each user is calculated for all products. Now, based on Equation (4), we calculate the MAE criterion, which is the average error of all users for all recommendations. The MAE criterion in this research is equal to:

\[
MAE = 0.2620\%
\]

This criterion shows the extent of the predicted ranking deviation from the actual ranking of users. Therefore, the accuracy value of the proposed method is as follows:

\[
Accuracy = (1 - MAE) \times 100 = 99.7380\%.
\]

4.4. Performance Evaluation of the Proposed Method Based on Product Recommendations. Recommendation criteria in medical advice methods help physicians to select the highest quality treatment options available outside the collection. These criteria are defined as the ratio of useful advice given to patients among all recommendations. Popular criteria in this area include inverse rate, weight errors, receiver-operating characteristics (ROC) and precise reminder curve (PRC), accuracy, sensitivity, and \( F \)-criteria. The most well-known criteria used in most research are accuracy and sensitivity. These two criteria are defined as follows [50]:

\[
\text{Accuracy} = \frac{\text{Correctly recommended items}}{\text{Total recommended items}},
\]

\[
\text{Recall} = \frac{\text{Correctly recommended items}}{\text{Total useful recommended items}},
\]

\[
\text{Precision} = \frac{\text{Correctly recommended items}}{\text{Total recommended items}}.
\]
In general, the criterion of accuracy is defined as the ratio of items related to real items, and the criterion of sensitivity is defined as the ratio of appropriate items among all relevant items. However, in this study, the accuracy criterion is defined as the ratio of the number of correct recommendations to the total number of recommendations, and the sensitivity criterion is defined as the ratio of the number of correct recommendations to the total number of useful recommendations. Performance evaluation criteria of the $F$-criterion is defined to help simplify the accuracy and reminder criteria in a single criterion. The value obtained makes the comparison between algorithms and the whole data very simple and easy. The measurement relationship of $F$-measure is defined as follows:

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (11)$$

Accordingly, according to the relationships, Table 4 shows the values related to the evaluation criteria.

As shown in Table 4, the proposed method, which is a combination of the nearest neighbor algorithm to classify the type of chronic disease and a participatory filter to provide the necessary treatment recommendations for the disease, in terms of evaluation criteria, good performance relative to the class, it has NN, SVM, and NB clauses. The high accuracy of the proposed method indicates the high ability of this method in teaching the model based on the nearest neighbor method in the proposed method. Accordingly, the accuracy diagram of the proposed method for 450 test samples using classification methods is shown in Figure 4.

As shown in Figure 4, the proposed method, using a combination of the nearest neighbor method and the participatory filter, has performed well in terms of accuracy in the chronic patient dataset. Another criterion that has been evaluated in the proposed method is the criterion of sensitivity, which is useful as a percentage of the correct recommendations provided to the total recommendations provided in the proposed method. Figure 5 shows the diagram of the sensitivity criterion in the proposed method.

As shown in Figure 5, in the combination of the nearest neighbor method and the participatory filter, more than 90% of the recommendations given to patients are correct, and this shows the high capability of the proposed recommender method. Another criterion used in the proposed method is accuracy, which is the percentage of correct recommendations given to the total recommended recommendations. Figure 6 shows the correctness diagram of the proposed method.

As shown in Figure 6, in the combination of the nearest neighbor method and the participatory filter, about 89% of the recommendations given are correct. Another criterion used in the proposed method is $F$-measure, which is a combination of two criteria of accuracy and sensitivity. Figure 7 shows the $F$-measure diagram of the proposed method.

As shown in Figure 7, the $F$-measure criterion in the combination of the nearest neighbor method and the participatory filter is high and is about 90%.

4.5. Comparison of the Proposed Method with Previous Methods. After evaluating the proposed method, to evaluate the validity of the proposed method, we compare it with previous methods in this field. As shown in Tables 2–4, the performance accuracy of the proposed method is determined on the dataset of patients with chronic diseases and related charts and evaluation criteria. The proposed method can now be compared with previous methods [51, 52] in the same dataset. Therefore, Figure 8 shows a comparison of
the proposed method with previous methods in predicting the class label of chronic diseases.

As shown in Figure 8, the proposed method has a higher accuracy in diagnosing patients than other previous methods due to the use of the method of close neighbors.

5. Conclusion

Chronic disease prediction systems are very useful as an aid to the health and care of the sick community and thus reduce mortality even in young and middle-aged people. The chronic disease prognosis system can be an important issue in medicine, helping ordinary people to be aware of their health status, because normal people are late for health check-ups. People who do this periodically do not have the necessary information about their health status. By having the results of the chronic disease prediction system, if you have these diseases, you can be aware of the progression of the disease and prevent the progression of the disease in the early stages. Then, based on different parameters, appropriate treatment recommendations can be provided to patients. The accuracy of the diagnosis of chronic disease prediction systems depends on the accuracy of the proposed model in determining diagnostic patterns. Therefore, the more accurate the model, the greater the ability to predict patients at risk for chronic disease. In this study, a method for diagnosing chronic diseases based on patients’ health records using a referral system is presented. In this research, after extracting the basic features, the nearest neighbor method has been used. Experimental results show that the proposed method has a good performance in terms of evaluation criteria. The high accuracy of the proposed method indicates the high ability of this method to provide appropriate treatment recommendations in the proposed method. The results of the experiments in the proposed method show that, firstly, this method is 96.5% accurate in diagnosing the type of chronic patient, which is a very good value, and compared to the best available methods that were implemented based on neural networks, about 9% has improved. Second, the MAE rate of the proposed method in finding similar patients in the community of specific chronic patients was only 0.25%. The proposed method owes these appropriate values for accurate diagnosis of disease type and error of identification of similar patients to the proper use of machine learning methods. A major limitation of the proposed method is the lack of access to real hospital data. Due to patients’ privacy concerns, medical centers do not easily provide real information to researchers, and therefore, the proposed method uses the open access data in the data repository. However, despite the real data, the improvement of the proposed method can be seen in practice.

In order to provide suggestions for future work, a combination of supervised or unsupervised learning methods with metaheuristics feature selection methods in order to finding useful features of patients to increase disease diagnosis accuracy can be advised. On the other hand, due to the fact that labeling in real data is a difficult task and most of the data related to patients in healthcare systems are unlabeled, the use of semisupervised methods to diagnose the type of disease in recommender systems can be helpful.

Data Availability

The data used to support the findings of this study are available in https://physionet.org/.

Conflicts of Interest

The authors declared that they have no conflicts of interest.

References


