Preliminary Medical Diagnosis Using Voice-Based Urdu Language Interface

Muhammad Sajid[1]*, Kamran Ahsan [2], Muhammad Khalid Shaikh[2]

[1]Department of Computer Science, University of Balochistan, Quetta.

[2]Department of Computer Science, Federal Urdu University for Arts, Science and Technology Karachi.

***Correspondence**: m.khalid.shaikh@fuuast.edu.pk

Citation | Sajid. M, Ahsan. K, Shaikh. M. K, "Preliminary Medical Diagnostic Using Voice-Based Urdu Language Interface", IJIST, Vol. 07, Issue. 04 pp 2819-2841, November 2025

Received | October 11, 2025 **Revised |** November 08, 2025 **Accepted |** November 15, 2025 **Published |** November 22, 2025.

Expert knowledge is stored in the knowledge base through an externalizing process in the form of facts, procedures, heuristics, and rules. The knowledge base helps to refine the present knowledge and insert new knowledge without recompiling a program. Medical diagnosis is one of the first knowledge-based areas in which expert system principles are applied. Almost all knowledge-based medical diagnostic systems take input symptoms in the form of text and rely on the English language. This is a hindrance to illiterate and non-native English speakers of developing countries to utilize the system, and unfortunately, Pakistan is one of them. In this connection, this paper proposed an indexing method for integrating the medical diagnostic knowledge base with a Pakistani National Language-based voice-oriented user interface for accommodating the illiterate.

Keywords: Diagnostic Tree, Knowledge Base, Medical Diagnosis System, Artificial Intelligence, Speech Recognition

Introduction:

The synthesis of experiences, context, interpretation, and reflection with information generates knowledge. Knowledge represents a refined form of information that exists cognitively within individuals; however, it is inherently more complex and costly to communicate than raw information. Knowledge can be categorized into two types: tacit and explicit [1][2]. Tacit knowledge resides within individuals [3] and manifests through expert skills, practical experience, common sense, and sound judgment. In contrast, explicit knowledge is found in the form of procedures, knowledge base, and protocols [4][5][6][7]. Tacit knowledge can further be divided into two dimensions: technical and intellectual knowledge [8][9]. Technical knowledge is informal and associated with "know-how," making it difficult to articulate, whereas intellectual knowledge relates to one's beliefs, ideas, and values. Quality diagnostic services and healthcare cannot be provided without tacit knowledge. Many giant healthcare organizations provide opportunities for their doctors to share experiences, judgment, and skills to improve their knowledge. These opportunities are provided by organizing conferences, group discussions, and constituting a medical expert panel. Group discussion and expert panels enable practitioners to take collective decisions on a particular medical problem. However, access to specialists remains limited, especially in developing nations. The scarcity of medical professionals and the lack of adequate healthcare infrastructure have led to significant health crises in these regions.

The severity of these problems can be mitigated by implementing a knowledge-based expert system.

The Knowledge base is a centralized repository of data or information that comprises experts' tacit knowledge by capturing experiences, judgment, and skill through the externalizing process, as shown in Figure 1.

The knowledge-based system in medical diagnosis was taken up in the 1970s. INTERNIST is one of the first knowledge-based clinical support systems developed by Myers in 1974 at the University of Pittsburgh [10].

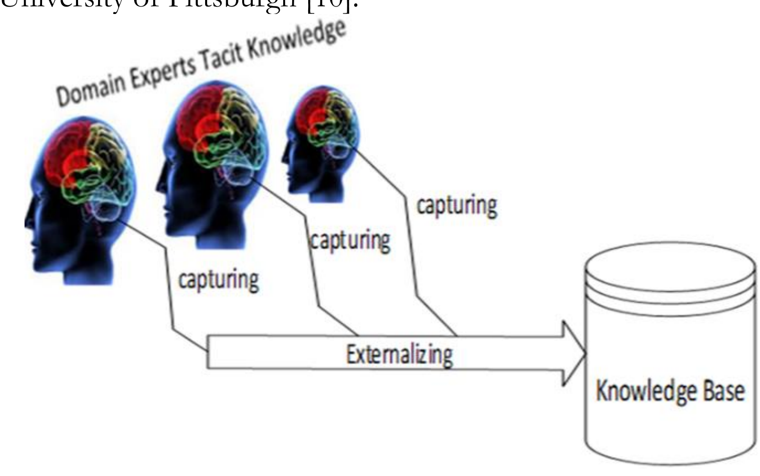

Figure 1. Knowledge capturing and externalizing

Knowledge generated through human interactions in the digital environment, combined with insights into their device preferences, has the potential to enhance the training of artificial intelligence systems. Such systems can be leveraged to strengthen global health initiatives by providing predictive, diagnostic, and recommendation capabilities, especially for underserved communities [11]. It is observed that in the last decade, the use of knowledge base systems has been embedded with other technologies, such as mobile and speech recognition technologies. It also observed that the large majority of knowledge base medical-diagnosis systems are based on the English language. As a result, a large portion of the illiterate population and those lacking proficiency in English are unable to effectively use these systems.

In this context, this paper proposed an indexing technique for integrating the knowledge-based medical diagnostic with a Pakistani regional language-based user interface, since can be more accessible to people who do not speak English or have limited English proficiency.

This is particularly crucial in regions with significant language barriers or where most individuals do not speak English as their primary language. People are more likely to engage with a system presented in their native language, especially if they are not proficient or comfortable with English.

This can result in greater user adoption and engagement. Incorporating national languages such as Urdu, Hindi, Malay, Bahasa, Javanese, or Bengali, as well as regional languages like Pashto, Punjabi, Marathi, or Hindko, can help users better understand and interact with the system. This approach enhances accuracy and efficiency while fostering improved communication and collaboration among users.

Form-based interaction that relies on "clicks" and "selections" enables users to submit their symptoms more efficiently and with fewer errors. Moreover, it allows medical staff to interpret and process the information more easily and accurately.

It is important to note that implementing a local language-based MIS can also have some challenges, such as the need for translation and localization, as well as the potential for language-based barriers to collaboration among multilingual teams.

Objectives:

An objective of this research is to design and develop a novel voice-based intelligent medical diagnosis system. The purpose of this system is to enable users to communicate symptoms in their local language. This can improve accessibility for individuals who are not proficient in English. Another objective is to incorporate multiple Pakistani regional languages into the diagnostic interface to ensure inclusivity and usability for rural populations belonging from various localities of Pakistan. Yet another objective is to construct a practitioner-validated knowledge base of symptoms in Roman Urdu, using expert medical knowledge and reliable clinical resources, so as to ensure accurate mapping between voice inputs and medical conditions. Finally, this research is also aimed at developing a customized diagnostic tree capable of predicting diseases.

Novelty Statement:

This research is novel in the sense that it has introduced the first Urdu and regional-language voice-based medical diagnosis system, which integrates a practitioner-validated knowledge base with a custom diagnostic tree, enabling accurate, speech-driven symptom interpretation. Even the illiterate and rural populations can use this application without difficulty.

Significance of the Current Research:

This Urdu supported voice-based medical information systems can provide a number of benefits to local healthcare dispensaries, organizations, and professionals, as well as to patients. Some of the potential industrial benefits of such systems include enhanced patient care, improved diagnostic accuracy, greater administrative efficiency, reduced patient turnaround time, better communication, and increased patient satisfaction. Such systems help to reduce the risk of errors and increase the speed of data entry, allowing healthcare professionals to focus on providing care to patients rather than on administrative tasks. They also provide easy access to important medical information, allowing healthcare professionals to make more informed decisions about patient care. They streamline processes and reduce the time and effort required for tasks such as documentation, freeing up healthcare professionals to focus on other important tasks. As part of this, these systems can facilitate communication between healthcare professionals, helping to ensure that all team members have access to the same information and can collaborate more effectively and provide patients with easy access to their medical records and other important information, helping to improve their overall experience with the healthcare system.

Like many developing nations, countries such as Pakistan, India, Malaysia, Indonesia, Nepal, and Bangladesh face major challenges in the healthcare sector. According to the World Health Organization (WHO), several key standards must be met to ensure the provision of universally accessible healthcare services.

Healthcare services are universally accessible for everyone.

Adequate healthcare infrastructure, such as medical centers, hospitals, primary health services, and medical professionals, must be available for everyone.

All healthcare providers must respect dignity and deliver appropriate care, observing due consideration of cultural norms.

All healthcare services must be delivered promptly.

These standards include the availability of adequate healthcare infrastructure and personnel, the respect of cultural norms and dignity in care, and the timely delivery of services. However, in Pakistan, these standards are often not met due to a variety of factors, including shortages of medical staff and infrastructure, long distances to healthcare centers, cultural values, political instability, and a lack of a holistic approach to healthcare.

Many countries have successfully leveraged technology—such as medical assistant software and customer relationship management systems—to deliver healthcare services in regions facing medical infrastructure and accessibility challenges. However, in the case of Pakistan, while the technology and software are available at no cost, people are unable to use them because their use of English as the primary language is English, which presents a barrier for many individuals.

One major barrier to healthcare access is the fact that much of the medical technology and software available relies on English as the primary language. This presents a significant challenge for those with limited literacy or proficiency in English, which is a significant portion of the population. In fact, only about 11% of Pakistanis can understand English, and the majority of these individuals live in urban areas. More than 80% of Pakistanis cannot read or write in English, meaning that existing medical applications may not be accessible or useful to the majority of the population.

In order to address this issue, it is necessary to develop expert systems that are able to translate the knowledge base into regional languages and provide medical facilities in a language that is easily understood by the majority of the population. This can be accomplished by integrating speech recognition and mobile technologies with medical diagnostic applications.

With speech-based user interfaces, patients can describe their symptoms in their native language, making the process more accessible. Additionally, GPS technology allows for the tracking of doctors and ambulances, improving coordination and the overall efficiency of healthcare delivery. Our research suggests that an expert system that addresses multiple areas of medical care, including doctor availability, ambulance management, medical guidance, and medical diagnosis, could be beneficial in countries where English is not the first language. Previous research and literature review indicate that most medical applications use an English language-based user interface, with some using a graphical user interface (GUI) for inputting symptoms and others using SMS-based symptom input. However, none of the existing systems utilize speech-based user interfaces for symptom input or GPS for tracking patients, doctors, and ambulances. While these systems have been successful in improving healthcare services in other developing countries, they have not been as effective in underdeveloped countries due to the high levels of illiteracy and low proficiency in English among the population. The integration of speech recognition technology and mobile technology with medical diagnosis applications could potentially address these issues and be beneficial for people in underdeveloped Asian countries. Additionally, the existing medical diagnostic systems do not clearly depict the relationships among all components, making it difficult to understand their functions and interactions. Enterprise architecture could potentially improve the depiction of the functions and performance of all related components, and the integration of mobile phone GPS with medical diagnosis applications could enhance the delivery of healthcare by determining the nearest available doctor through tracking patient and doctor location.

Materials and Methods:

Medical diagnosis is a complex process that requires extensive experience and expertise. Since medical specialists are not always available, medical software can play a crucial role by assisting inexperienced doctors and non-professionals in diagnosing diseases more accurately and efficiently. All developing countries takes the advantage of these expert systems and facilitate their nation. Medical software plays an important role in providing quality medical services and handling multiple patients. Since mobile phone usage is widespread across both rural and urban areas, it is rare to find someone without access to one. Therefore, medical software designed to be mobile-friendly ensures easy accessibility and usability for all types of users. These systems are known as expert systems. An expert system is a computer program that operates within a specific domain, utilizing specialized knowledge typically held by human experts [12][13]. Numerous medical software programs are effectively serving in the medical area, and some are briefly discussed below.

Mycin:

MYCIN was developed by Dr. Edward Feigenbaum and Buchanan as the first system which is based on a Rule-Based medical diagnosis system at Stanford University in 1970 [14][15][16]. It consists of approximately 450 rules, and its generalized version, known as EMYCIN, is utilized for diagnosing blood infections [17][18]. It performed at nearly the same level of competence as human specialists in diagnosing blood infections and even surpassed general practitioners, offering significant support to the medical field.

MobDoc:

MobDoc [19] is a rule-based expert system that can be deployed in mobile devices in order to provide primary health care services in the absence of a doctor. This system is particularly beneficial in rural areas where healthcare infrastructure is lacking and paramedical staff are scarce. It collects patient symptoms through a question-and-answer format, where users select responses from given options, as illustrated in Figure 2. Using MobDoc is simple and efficient, as it can be installed directly on users' mobile devices, allowing access anytime and anywhere during a medical emergency. Users do not need to worry about how to interact with the system, as it provides guided options and user-friendly assistance throughout the process.

Figure 2. Screenshot of MobDoc [19]

An Adoptive Medical Diagnosis System:

AMDS is a rule-based expert system model developed for common diseases like Typhus, Malaria, Plague, and Typhoid, etc. [20][21]. The AMDS system, based on expert system (ES) technology, is highly beneficial for patients suffering from common diseases, as it can generate prescriptions similar to those provided by medical professionals. This system is particularly valuable in rural areas with limited or inexperienced medical staff. AMDS assists young doctors and paramedical personnel in the disease diagnosis process, as illustrated in Figure 3.

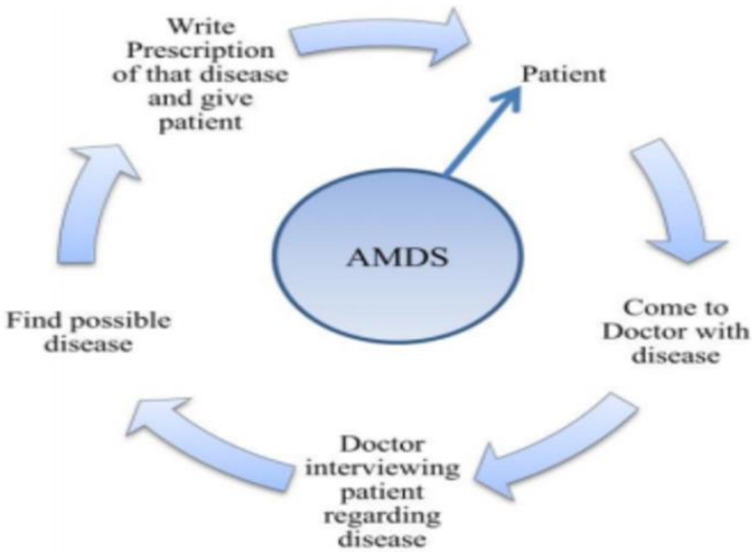

Figure 3. Adoptive Medical Diagnosis System [20].

Diagnosis Expert System:

DExS is built on a pattern-matching algorithm. To accumulate factual knowledge, data are collected concerning the association between signs and symptoms associated with patients. The signs, symptoms, and test reports are the determining factors of a particular disease. The main purpose of the DExS system is to connect the evidence (input data) with appropriate rules through the rule base. DExS is designed for professionals; it provides consultation, quick diagnosis, and justifications of the obtained results [22][23]. The working model of DExS is shown in Figure 4.

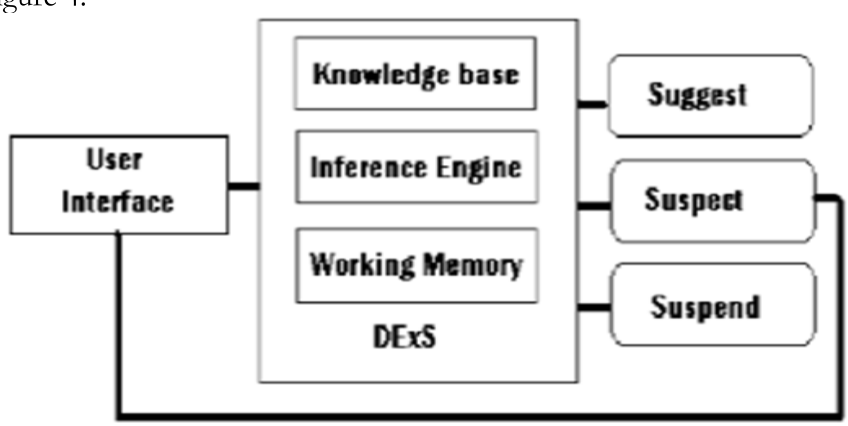

Figure 4. Working Model of DEX

Automated Mobile Asthma Monitoring:

The primary aim of this system design is to create a real-time application that can monitor and detect early signs of asthma during physical activity. This system promptly alert user upon detecting any irregularities to prevent a worsening of the condition. Additionally, the system provides users with feedback regarding their health status. To accomplish this, Uwaoma and Mansingh incorporated advanced algorithms capable of analyzing and classifying breath sounds, as well as identifying different levels of physical activity. These algorithms are executed on a smartphone, providing a practical and accessible platform for this application. [24].

Treatment of Hypertension in Pregnancy:

Hypertensive disorders represent a major cause of pregnancy-related maternal mortality worldwide. Similar to the non-pregnant population, hypertension is the most common medical disorder encountered during pregnancy and is estimated to occur in about 6–8% of pregnancies [25].

This system is built to address the shortage of obstetrician experts at the Reproductive Health Division in Kenya. This system assists inexperienced practitioners in the diagnosis and treatment of hypertension in Pregnancy. The knowledge base of this system includes facts about preeclampsia and hypertension ([26]. The Waterman [27] approach is followed to develop this system. The structure of the treatment of hypertension in the pregnancy expert system is shown in Figure 5.

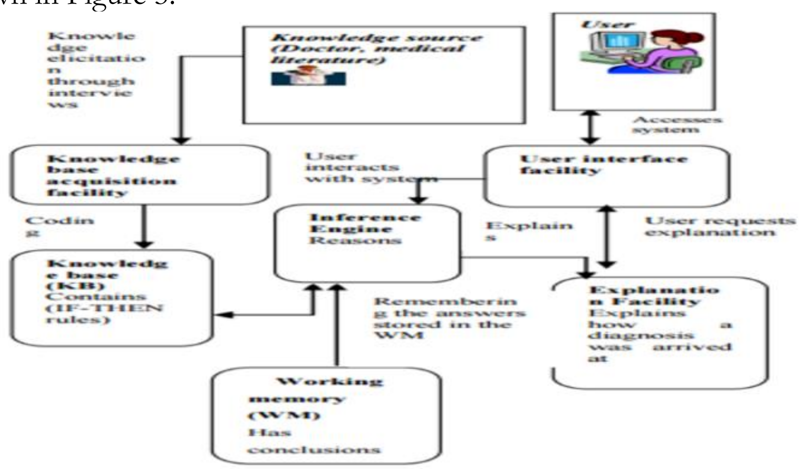

Figure 5. Treatment of hypertension in pregnancy expert system structure [26].

Neuro Fuzzy Expert System:

This system, developed for heart disease diagnosis, is based on genetic algorithms, fuzzy rules, and neural networks. It allows users to carry out diagnostic procedures using a reduced number of tests by leveraging the optimization capabilities of genetic algorithms. The fuzzy rule and neural networks enable the system to perform prediction effectively [28]. The working model of the Neuro fuzzy expert system is shown in Figure 6.

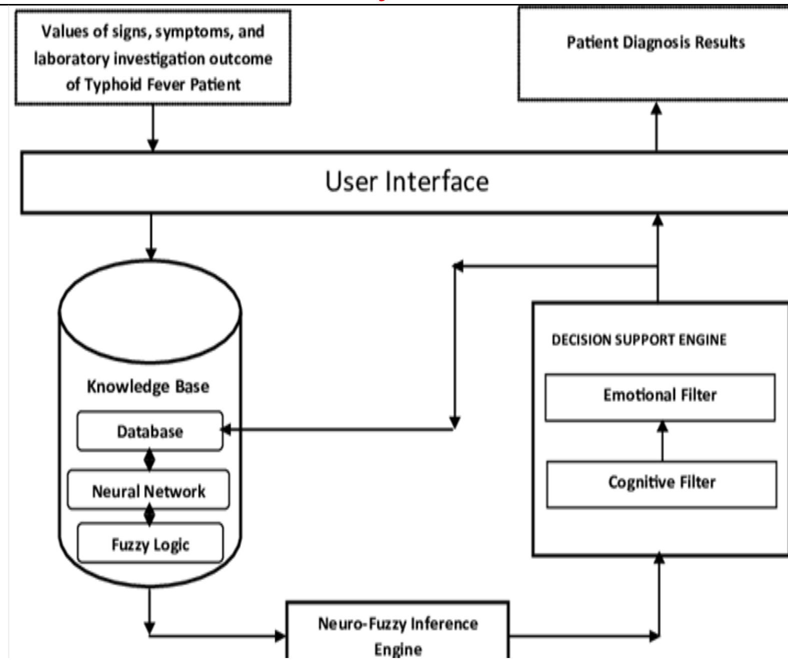

Figure 6. Neuro-fuzzy system

Rule-Based Expert System for Neurological Disorders:

This system is capable of diagnosing 10 types of neurological diseases by employing a rule-based technique. Microsoft Visual Studio .NET and Prolog were used to build conceptual system models of this expert system. The Rule-Based Expert System for Neurological Disorders empowers the neurologists, neurology students, and patients to get guidance or advice about neurological disorders by putting in the symptoms of illness [29].

In this system, the user was presented with a set of queries in the interactive window, for which the user had the option to answer yes or no. The set of questions was prepared according to the symptoms shown by patients with neuromuscular diseases. Based on the feedback given by the patient, the RETE algorithm searched the knowledge base for possible pattern matches. If there was a rule in the knowledge base that matched the patient's symptoms, the system displayed the possible diagnosis in the recommendation window.

Expert System for the Management of Hypertension:

This is a web-based Hypertension Management Expert System developed using fuzzy logic techniques, as illustrated in Figure 7. Expert System for the Management of Hypertension takes input in the form of blood pressure reading, body-mass-index, and age. The fuzzy-rules compute these parameters and produce results in the form of hypertension risk. The web-based interface is designed using PHP and HTML, MySQL is used for managing the knowledge base, and an Apache server is used for server-side scripting [30].

Figure 7. Expert System for the Management of Hypertension [30].

Third-world countries are surrounded by healthcare and education problems due to a shortage of professionals and infrastructure. The mortality rate is very high in these countries, especially in rural areas, due to the unavailability of medical professionals. The knowledge-based medical software can fill the gap in the shortage of medical professionals and may reduce the mortality rate. Many developing countries are taking advantage of medical software and improving healthcare services. Pakistan is also developing and facing serious healthcare problems. Pakistan can't take the assistance of knowledge-based medical diagnostic software because of illiteracy. The literacy rate of Pakistan is very low, especially in rural areas. Most Pakistani people can't understand English, which is why they are unable to use medical diagnostic software. In this connection, this research proposed a new method (indexing method) for integrating the knowledge base medical diagnostic system with a speech-based user interface in the regional languages of Pakistan.

Methodology:

This method comprised two phases: the first involved developing a knowledge base and diagnostic tree, while the second focused on the diagnostic process.

Phase 1: Development of a knowledge base diagnostic tree:

The development of a knowledge base was the first phase of the indexing method. Three major tasks were carried out in this phase: (1) developing a knowledge base for symptoms, (2) developing the diagnostic tree, and (3) building a knowledge base for diagnostic results and prescriptions based on the diagnostic tree.

Task 1: Development of a knowledge base for symptoms:

The symptom knowledge was acquired from practitioners and stored in the knowledge base along with the Symptom Identification Number (SymID), Symptom Name in English (SymNameE), and Symptom Name in Roman Urdu (SymNameU), as shown in Table 1.

Task 2 (Development of Diagnostic Tree):

The second step of Phase 1 involved developing the diagnostic tree based on the collected symptom data, as illustrated in Figure 8. This diagnostic tree was constructed using Symptom Identification Numbers (SymIDs), developed in consultation with medical practitioners and by referencing online diagnostic platforms such as WebMD and similar resources.

Table 1. Symptoms Knowledge base

Symid	SymNameE	SymNameU
1	Sore Throat	Galay Kharab
2	Fever	Bukhar
3	Headache	Sar dard
4	Sneezing	Cheenkna
5	Discharge Nose	Naak Behna
6	Congested Nose	Naak Band
7	Cough	Khaansi
8	Mucus	Balgam
9	Itchy Eyes	Aankh Dard

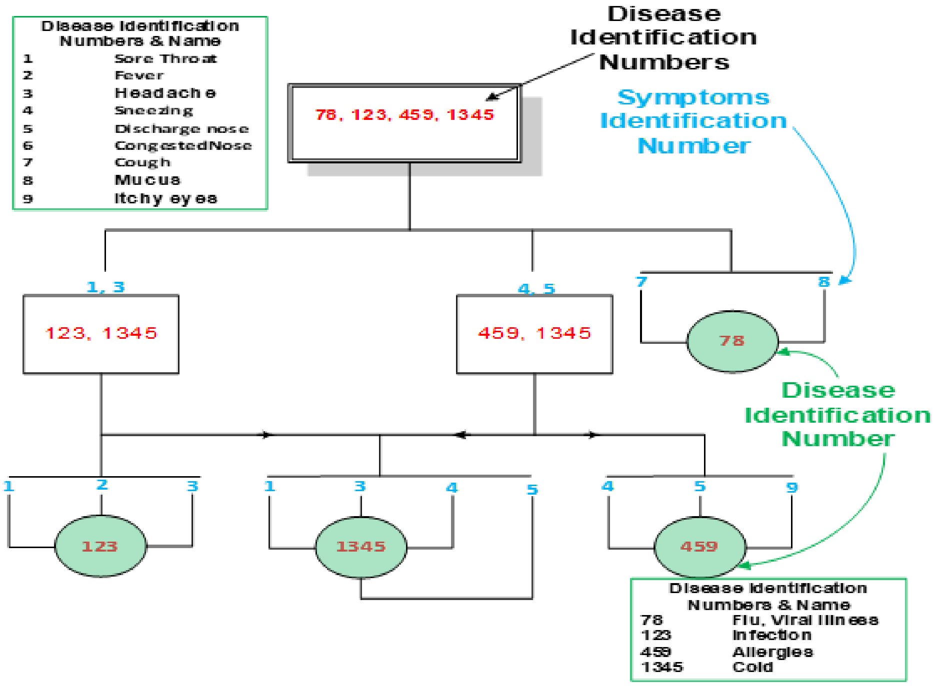

Figure 8. Diagnostic tree

Task 3 (Diagnostic & Prescription Knowledge Base):

The final task of Phase 1 was to build a diagnostic and prescription knowledge base by considering the diagnostic tree and consulting physicians, as well as verified online material, as shown in Table 2. This knowledge base contained information about disease treatment in the form of prescriptions and medications.

Table 2. Diagnostic & prescription knowledge base

Diagnosis		
DiagID	**Disease**	**Prescription**
78	Flu, Viral illness	oseltamivir Tamiflu 75mg (twice a day for 5 days)
123	Infection	Augmentin 375 mg(1 Tablet three time for 3 to 5 days)
456	Allergies	Softin Tab 10 mg (one a day) for up to 5 days
1345	Cold	Arinac Tab 200mg (one Tablet Three Time a Day) for 3 to 5 days.

Phase 2 (Diagnostic Process):

The second phase of the indexing method consisted of six steps, including speech grammar, symptom collection, arranging symptoms in an array, obtaining symptom IDs, sorting symptom IDs, and acquiring the diagnostic result, as illustrated in Figure 9.

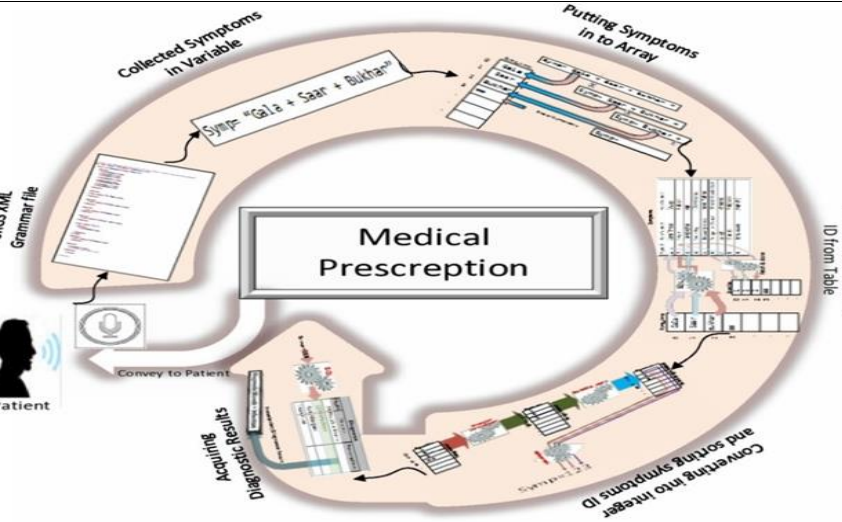

Figure 9. Diagnostic process

Speech Grammar:

The speech-based user interface gathered symptoms from the user/patient in the Urdu language by using a specially designed Roman Urdu grammar, as shown in Grammar-XML 1. This grammar was developed in Extensible Markup Language (XML) and was called the Speech Recognition Grammar Specification (SRGS). SRGS grammars enabled the construction of complex voice-based user interfaces by providing features such as specifying the order of words and phrases, merging words from multiple lists, assigning weights to words or phrases, and using special rules for speech recognition. Figure 10 shows a sample of the SRGS XML file for speech recognition. This file describes the order of words and the rules; for example, the <one-of> and <item repeat="0-1"> tags were used for defining the rules for selecting words from an item list. <one-of> meant selecting only one word from the list, while <item repeat="0-1"> meant that the word could be recognized zero or one time [31].

```
<grammar root="phoneRule" version="1.0"
xmlns="http://www.w3.org/2001/06/grammar"
xml:lang="en-US" tag-format="semantics/1.0">
  <rule id="phoneRule" scope="public">
    <one-of>
      <item>khaansi</item>
      <item>Saar</item>
       <item>Sir</item>
      <item>Gala</item>
       <item>Gulay</item>
      <item>Gaalay </item>
      <item>jism</item>
      <item>Aankhay</item>
      <item>Naak</item>
      <item>Bukhar</item>
       <item>Booker</item>
      <item>Cheenkana</item>
      <item>naak Behna</item>
      <item>Balgham</item>
    </one-of>
      </rule>
  </grammar>
```

Grammar-XML 1

Symptoms Collection:

The user interface collected symptoms one by one from the patient according to the defined rules of the SRGS XML specification and added them into a string variable. This string variable held only the first word of the symptom phrase. For example, when the user/patient said the symptom phrase "saar maa dard ho raa hay," the string variable stored the first word of the phrase, "saar." The "+" sign was used as a separator between two or more symptoms. Figure 10 shows the collected symptoms in the string variable Symp [32] .

```
Symp= "Gala + Saar + Bukhar"
```

Figure 10. Collected symptoms separated by the "+" sign

Arrange Symptoms in the Array:

After collecting the symptoms, the next step involved dividing them so that each symptom was stored individually in a separate string array referred to as Symp_Array. This array was later used for obtaining symptom IDs from the symptom table [33]. The splitting and transferring of symptoms from the string variable Symp into the array were performed by employing the algorithm described in Algorithm No. 1.

Algorithm/Pseudocode No.1:

Here, Symp is a string variable to hold all symptoms, Symp_Array is a String array to Hold Individual Symptoms, and LOP is a counter variable.
1. [Initialize counter] Set LOP:-0
2. Repeat Steps 3 to 5 while LENGTH(Symp)>1 [Repeat until variable Symp Length gather than one]
3. Symp_Array[LOP]:=SUBDTRING(Symp,0,(INDEX(Symp,'+')-1)) [Employing substring function for getting symptoms from variable Symp by determining the location of the separate mark sign "+"]
4. Symp:=DELETE(Symp,0,(INDEX(Symp,'+')+2)) [Removing Symptoms one by one from variable Symp]
5. Set LOP:=Lop+1 [Increment in Counter variable]
   [End of Step 2 Loop]
6. Symp_Array[Lop]:="##" [Marking end of Symptoms in Symp_array]
7. Exit

Algorithm No. 1:

Algorithm/Pseudocode No. 1 treated Symp as a string variable holding all symptoms, Symp_Array as a string array holding individual symptoms, and LOP as a counter variable [34].

The algorithm showed the results of each iteration: in each iteration, one symptom was added to the array and then removed from the string variable. This removal reduced the length of the variable. When the length of the array became 1 or 0, the loop terminated, and the array was marked with the end-of-symptoms indicator "##," as illustrated in Figure 11.

Getting symptom ID:

After completing the splitting and storing of symptoms into the array, the next activity was to fetch the Symptom ID from the symptom database table and store it in the SympNo array. This activity was performed by the algorithm described in Algorithm No. 2.

Algorithm/Pseudocode No.2:

Here, Symno is a string array to hold Symptom ID, LOP is a counter variable, Symp_Array is also a string Array holding Symptoms, SmSQL is a string variable to hold SQL Query.
1. [Initialize counter] Set LOP:-0
2. Repeat Step 3 and 4 While (Symp_Array[LOP]!="##") [Repeat until headcounter the "##" sign]

3. (a) SymSQL="Select Symid from Symptoms where SymNameU=" + Symp_Array[LOP] + " " [Putting SQL query in variable SymSQL]
   (b) System.Data.OleDb.OleDbDataAdapter(SymSql, Connection) [Passing Query through SymSQL variable to database connection]
   (c) Symno[LOP]=Tables["Semptoms".Row[0][0].ToString() [Fetching and holding Symptom ID symptom table in Symno Array]
4. Set LOP:=LOP+1 [increment in counter variable]
   End of step 2 loop
5. Set Symno=[LOP]:="##" [Mark End of Symptom]
6. Exit

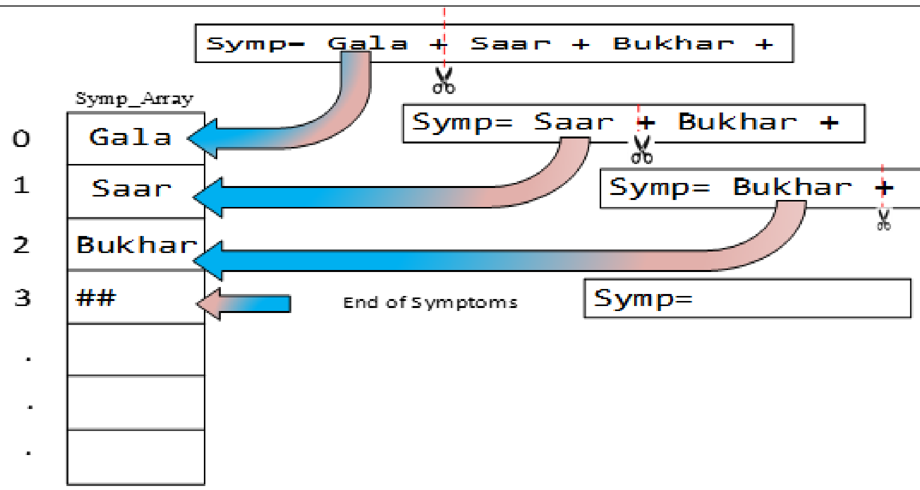

Figure 11. Results in each iteration of an algorithm/pseudocode No.1

Algorithm No.2 for Getting Symptoms ID from the Table:

The algorithm demonstrated all steps for obtaining symptom IDs from the symptoms table and storing them in the SympNo array. The function took one symptom at a time from Symp_Array until the counter reached the "##" symbol. Each symptom was passed as a key to an SQL query through a database connection to fetch the SympID from the symptoms table. Finally, the SympID was stored in the SympNo array, as shown in Figure 12.

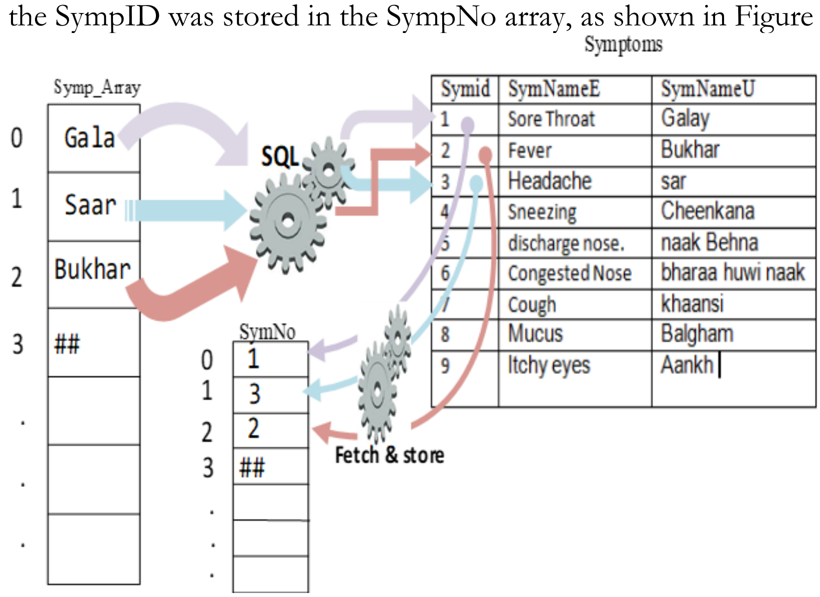

Figure 12. Getting symptoms not from the table

Sorting Symptom ID:

These Symptom ID numbers were stored in the SympNo array as strings. Before being used in the diagnostic process, they were passed through an integer conversion, sorting, and concatenation process by employing the algorithm shown in Algorithm No. 3.

Algorithm/Pseudocode No.3:

Here, LOP is a counter variable, Symint is an integer array to hold Symptom ID in the form of an integer value, Bubble sort is a procedure for sorting the array, Symp is a string variable to hold Symptom ID after Concatenation.

1. [Initialize counter] Set LOP:=0
2. Repeat Step 3 and 4 While(Symno[LOP]!="##" [Repeat until headcounter the ## sign]
3. Symint[LOP]=Convert.ToInt16(Symno[LOP]) [Convert each Symno in to integer and store in Symint Array]
4. Set LOP:=LOP+1 [Increment in Counter variable]
   [End of step 2 loop]
5. Bubble_Sort(Symint) [Apply Bullbe_Sort Procedure]
6. [Initialize Counter] Set LOP:=0
7. Repeat Step 3 and 4 while(LOP<Length(Symint)) [Looping]
8. SetSymp:=Symp // convert.ToString(Symno[LOP]) [Concatenate Symno into Symp variable]
9. Set LOP:=LOP+1 [increment in counter variable]
   [End of Step 7 loop]

10.EXIT

Algorithm No. 3. (Convert to integer, Sort and concatenate):

In the initial stage of Algorithm No. 3, the symptom IDs were converted from a string data type to an integer data type and stored within an integer array. The second phase sorted the integer array (SymInt) in ascending order. The third phase collected each sorted symptom ID from the array and combined them into a single variable using the concatenation procedure, as illustrated in Figure 13.

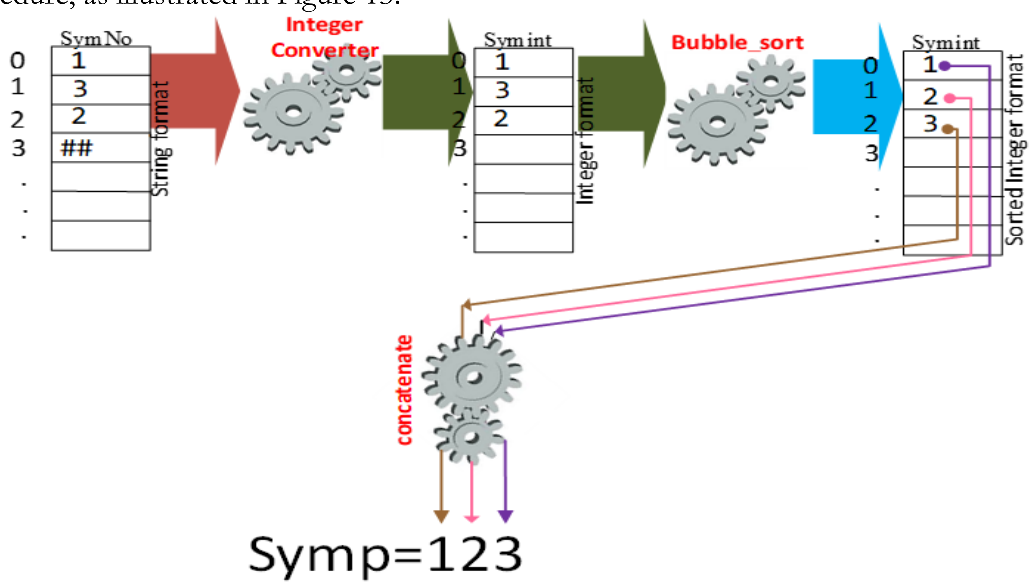

Figure 13. Phases of Algorithm No 3.

Acquire Diagnostic Result:

Finally, the concatenated form of symptom IDs stored in the Symp variable was used as a primary key to retrieve the corresponding disease diagnosis and medical prescription from the diagnosis table. The diagnostic result was obtained using the algorithm described in Algorithm No. 4.

Algorithm/Pseudocode No.4:

Here, DiagSQL holds SQL query, Symp holds Symptom ID, and DiagnosticResult holds Diagnosis Result, such as disease / Medical prescription. All variable data type is string.
1. DiagSQL:= "select * from Diagnosis where DigID=" + Symp + "" [Putting SQL query in variable DiagSQL]
2. System.Data.OleDb.OleDbDataAdapter(DiagSQL, Connection) [Passing Query through DiagSQL variable to database connection]
3. String DiagnosticResult=(DiagDs.Tables["Diagnosis"]. Rows[0][1]. ToString ()) [Fetching and holding disease / Medical prescription from the Diagnosis table in FinalResult variable]
4. Exit

Algorithm No. 4:

The first step of Algorithm No. 4 constructed a query by concatenating the Symp variable and storing it in the DiagSQL variable. In the second step, the DiagSQL variable was passed to SQL through a database connection to fetch the diagnostic result from the diagnosis table. Finally, the fetched information was stored in the variable diagnostic result, as shown in Figure 14.

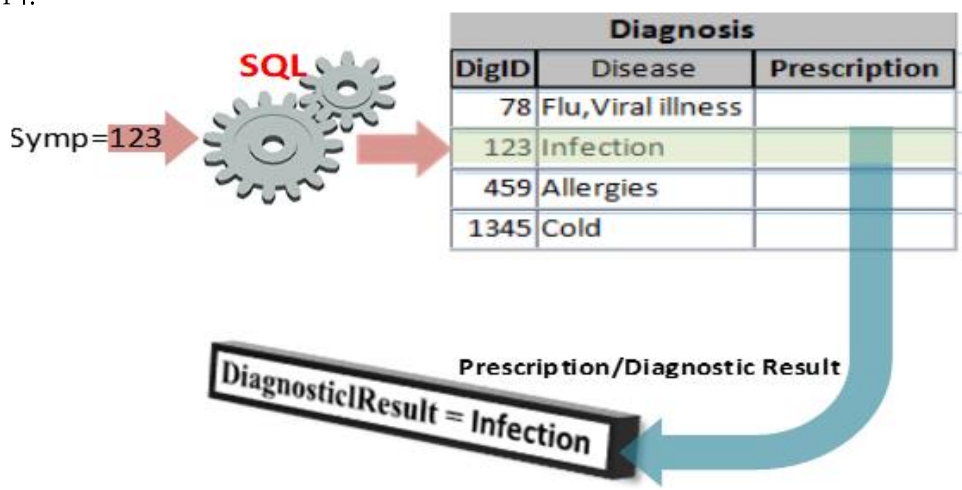

Figure 14. Getting Diagnostic information through algorithm no 4.

Result and Discussion:

The indexing method allowed the locals to benefit from medical diagnostic software by using speech-based user interfaces in their own regional languages. This method was able to integrate any Pakistani or Malaysian regional language (speech-based user interface) with the medical diagnostic system. The indexing method was initially evaluated in the Urdu language by attaching its speech grammar and a symptoms knowledge base containing information about symptom IDs, symptoms in English, and in Roman Urdu. The method took symptoms in the form of speech from the user/patient, converted them into the regional language, and performed the diagnostic process in six steps (already defined above) on the server, which then produced a prescription for the user/patient. The proposed method was particularly beneficial for individuals who were not proficient in English. It served remote populations by providing medical diagnostic services in areas where doctors or paramedical staff were unavailable. The indexing method bridged the gap between patients and primary medical services and provided diagnostic services at their doorsteps in their regional language.

The indexing method was evaluated in three phases: the first phase belonged to medical professionals, the second to IT/IS experts, and the third to users/patients.

Phase 1. (Medical Professionals):

Ten medical professionals were approached to evaluate the indexing method; five showed interest and evaluated it. The medical professionals evaluated the indexing method based on three key parameters: (1) the structure of the diagnostic tree, (2) the relationships between symptoms and diseases, and (3) the accuracy of diagnostic results and prescriptions.

Initially, the indexing method was applied to four types of illnesses, including Flu, Viral Illness, Infection, Allergies, and Cold. The medical professionals evaluated the illness diagnostic tree on the basis of these four types of diseases and assessed the decisions generated by the diagnostic tree. The medical professionals rated the diagnostic tree using the Likert scale, as shown in Table 3.

Table 3. Diagnostic tree assessment

Symptoms	Illness	Medical Expert	Strongly disagree	Disagree	Neither agree nor	Agree	Strongly agree
Cough (7), Mucus (8) 78	Flu, Viral illness	1				✓	
		2					✓
		3				✓	
		4				✓	
		5					✓
Sore Throat (1), Fever (2), Headache (3) 123	Infection	1					✓
		2					✓
		3					✓
		4				✓	
		5				✓	
Sneezing (4), Discharge nose (5), Congested Nose (6) 456	Allergies	1					✓
		2				✓	
		3				✓	
		4				✓	
		5				✓	
Sore Throat (1), Headache (3), Sneezing (4), Discharge nose (5) 1345	Cold	1					
		2				✓	
		3					✓
		4				✓	
		5					✓

According to Table 1, 55% of Medical Practitioners Agreed, 40% Strongly Agreed, and Only 5% Neither Agreed nor Disagreed with the Outcome of the Diagnostic Tree:

Medical experts assessed the relations between symptoms and disease by cross-relating them and rated them using a Likert scale, as shown in Table 4. Nearly 80% of medical experts strongly agreed that the provided symptoms were fully related to the above-mentioned diseases, while 20% agreed.

Table 4. Relations assessment

Symptoms	Illness	Medical Expert	Strongly disagree	Disagree	Neither agree nor	Agree	Strongly agree
Sore Throat, Fever, Headache, Sneezing, discharge from the nose, Congested	Flu, Viral illness, Infection,	1				✓	
		2					✓
		3					✓
		4					✓
		5					✓

Nose, Cough, Mucus, Itchy eyes	Allergies, Cold						

Parameter 3 (Diagnostic results and prescription):

Medical professionals evaluated the diagnostic results and prescriptions using a Likert scale, as presented in Table 5. It could be seen that 55% of medical professionals agreed with the diagnostic results and prescriptions, while 45% strongly agreed.

Table 5. Diagnostic results and prescription

DigID	Disease	Prescription	Medical Expert	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
78	Flu, Viral illness	Oseltamivir Tamiflu 75 mg (twice daily for 5 days)	1				✓	
			2				✓	
			3				✓	
			4				✓	
			5				✓	
123	Infection	Augmentin 375 mg (1 tablet three times a day) for 3 to 5 days	1				✓	
			2				✓	
			3					✓
			4					✓
			5					✓
456	Allergies	Softin Tab 10mg (Once a day) for 5 days	1				✓	
			2					✓
			3					✓
			4					✓
			5					✓
1345	Cold	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	1				✓	
			2				✓	
			3					✓
			4				✓	
			5					✓

Phase 2. (IT/IS experts):

During the evaluation, 10 IT/IS experts were approached to assess the indexing method; four declined, while six expressed interest. These experts evaluated the method based on three specific parameters. The first parameter is the method of symptom collection, the second is the integration of the various components of the indexing method, and the third involves evaluating the algorithms.

Parameter 1 (Symptoms Collection Method):

More than 66% of experts agreed that the symptoms collection method was user-friendly and comfortable for the Pakistani population because it used speech in the native language. About 34% of IT/IS experts were somewhat satisfied but argued that noise factors could have affected the performance of the speech-based input interface.

Parameter 2 (Integration of Different Components):

Almost 83% of IT/IS experts were satisfied with the integration of different components of the indexing method, including speech grammar, the knowledge base, arranging symptoms in the array, and obtaining and sorting symptom IDs.

Parameter 3 (Algorithm/Pseudocode):

These algorithms were identified as highly effective in addressing the specific problem domain by 66% of IT/IS experts, while 34% believed that there was room to further optimize these algorithms.

Phase 3 (Users/Patients Test):

To test the indexing method, 34 users were voluntarily selected. The users provided different symptoms (selected from a given list) in Urdu and received specific medical prescriptions. The test results were shown in Table 6. According to Table 6, the outcomes of the test results for the 20 users whose input speech was successfully recognized were 100% satisfactory and matched the expert opinions. However, for 14 users, the system did not properly recognize the input speech. The system was unable to recognize their speech due to the users' phonology and certain technical constraints.

The evaluation and test results are evidence that the indexing method is reliable and most effective for speech-based medical diagnostic applications in Pakistani regional languages. The experimental prototype test results show that the outcome of this method, which is a medical prescription, is 100% matched by the medical expert's evaluation results. However, the speech technology and noise may affect the performance of the indexing method.

Table 6. Test Results

Sno	Symptoms input	Results	Match with the Medical expert evaluation result in phase 1
1	Galay, sar, Cheenkana, naak Behna	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
2	Sore Throat, Fever, Headache	Augmentin 375 mg (1 tablet three times a day) for 3 to 5 days	✓
3	Sore Throat, Headache, Sneezing	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
4	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (Once a day) for 5 days	✓
5	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (One a day) for 5 days	✓
6	Sore Throat, Fever, Headache	Augmentin 375 mg (1 tablet three times a day) for 3 to 5 days	✓
7	Cough, Mucus	Oseltamivir Tamiflu 75 mg (twice daily for 5 days)	✓
8	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (Once a day) for 5 days	✓
9	Galay, sar, Cheenkana, naak Behna	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
10	Sore Throat, Headache, Sneezing	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓

11	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (One a day) for 5 days	✓
12	Cough, Mucus	Oseltamivir Tamiflu 75 mg (twice daily for 5 days)	✓
13	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (Once a day) for 5 days	✓
14	Sore Throat, Headache, Sneezing	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
15	Cough, Mucus	Oseltamivir Tamiflu 75 mg (twice daily for 5 days)	✓
16	Sore Throat, Fever, Headache	Augmentin 375 mg (1 tablet three times a day) for 3 to 5 days	✓
17	Sore Throat, Headache, Sneezing	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
18	Sneezing, Discharge from the nose, Congested Nose	Softin Tab 10mg (Once a day) for 5 days	✓
19	Galay, sar, Cheenkana, naak Behna	Arinac Tab 200mg/30mg (1 tablet three times a day) for 3 to 5 days	✓
20	Cough, Mucus	Oseltamivir Tamiflu 75 mg (twice daily for 5 days)	✓

Discussion:

The findings of this study demonstrated that the proposed indexing method effectively integrated Urdu and regional-language speech recognition with a knowledge-based diagnostic framework. The evaluation conducted with medical professionals confirmed that the diagnostic tree and symptom–disease mappings were clinically appropriate, and the prescriptions generated by the system closely aligned with expert judgment. This reinforces the reliability of the underlying knowledge base and the diagnostic logic developed for the system.

The responses from IT/IS experts indicated that the overall system architecture—particularly the combination of SRGS grammar, symptom parsing algorithms, and ID-based indexing—was technically sound and suitable for real-world deployment. Although some experts suggested further optimization of the algorithms, the system was generally assessed as efficient and correctly integrated across its functional components.

User testing further highlighted the practical relevance of the proposed approach. The system successfully recognized symptoms for the majority of participants and provided accurate diagnostic outcomes when speech was correctly interpreted. However, recognition performance was affected by accent and phonological variations, especially among rural users. This limitation reflects the broader challenges of speech technology in multilingual environments and indicates the need for more extensive language modeling and noise-handling mechanisms.

Overall, the discussion affirms that the indexing method can serve as a viable foundation for voice-based diagnostic applications in Pakistan and similar regions. It bridges a critical gap by enabling illiterate or non-English-speaking individuals to access basic diagnostic services in their native language. Nonetheless, to enhance system robustness and scalability, future work should focus on expanding the symptom set, refining the speech-recognition component, and deploying mobile-based real-time versions for wider adoption in rural and resource-constrained communities.

Conclusions:

The majority of existing medical diagnostic systems rely on the English language and use a text-oriented user interface for getting symptoms, which is why these systems are not effective in South Asian countries due to low literacy rates, a lack of technical skills, and an unawareness of English. The proposed method empowers the Pakistani nation to benefit from the knowledge-based medical diagnostic system through the voice-oriented user interface. The proposed method integrates the medical diagnostic knowledge base with a voice-oriented user interface that is based on Pakistani regional languages. Initially, the proposed method is evaluated on the Urdu language by integrating a speech-based user interface with a knowledge-based medical diagnostic system. Nonetheless, the speech recognition technology, noise, and phonology of users might influence the execution of the indexing method.

References:

[1] E. Borousan, "Evaluating factors that cause problem in implementation of knowledge management in Iran's oil and gas industry," *AFRICAN J. Bus. Manag.*, vol. 6, no. 34, Aug. 2012, doi: 10.5897/AJBM11.1446.

[2] I. J. K. Lindy M. Kregting, Nicolien T. van Ravesteyn, Wolfert Spijker, Tessa Dierks, Clare A. Aitken, H. Amarens Geuzinge, "Effects of a leaflet on breast cancer screening knowledge, explicit attitudes, and implicit associations," *Patient Educ. Couns.*, vol. 103, no. 12, pp. 2499–2507, 2020, doi: https://doi.org/10.1016/j.pec.2020.06.032.

[3] M. S. Adam and C. Urquhart, "IT capacity building in developing countries: A model of the Maldivian tourism sector," *Inf. Technol. Dev.*, vol. 13, no. 4, pp. 315–335, Oct. 2007, doi: 10.1002/ITDJ.20065.

[4] Y. N. Cheah and S. S. R. Abidi, "Augmenting knowledge-based medical systems with tacit healthcare expertise: Towards an intelligent tacit knowledge acquisition info-structure," *Proc. IEEE Symp. Comput. Med. Syst.*, pp. 264–269, 2001, doi: 10.1109/CBMS.2001.941731.

[5] Protocol for a scoping review to evaluate the extent of utilisation of healthcare services by asthma patients in sub-Saharan African countries, "Protocol for a scoping review to evaluate the extent of utilisation of healthcare services by asthma patients in sub-Saharan African countries," *BMJ Open*, vol. 11, no. 8, 2021, [Online]. Available: https://bmjopen.bmj.com/content/11/8/e046294

[6] D. S. Aine Carroll, "Use of complexity theory in health and social care: a scoping review protocol," *BMJ Open*, vol. 11, no. 7, p. e047633, 2021, [Online]. Available: https://bmjopen.bmj.com/content/11/7/e047633

[7] D. L. Hudson, M. E. Cohen, P. W. Banda, and M. S. Blois, "Medical Diagnosis and Treatment Plans Derived from a Hybrid Expert System," *Hybrid Archit. Intell. Syst.*, pp. 329–344, Sep. 2020, doi: 10.1201/9781003068075-17.

[8] T. Haldin-Herrgard, "Difficulties in diffusion of tacit knowledge in organizations," *J. Intellect. Cap.*, vol. 1, no. 4, pp. 357–365, Dec. 2000, doi: 10.1108/14691930010359252.

[9] I. Yazici, O. F. Beyca, O. F. Gurcan, H. Zaim, D. Delen, and S. Zaim, "A comparative analysis of machine learning techniques and fuzzy analytic hierarchy process to determine the tacit knowledge criteria," *Ann. Oper. Res. 2020 3081*, vol. 308, no. 1, pp. 753–776, Jun. 2020, doi: 10.1007/S10479-020-03697-3.

[10] D.A. Wolfram, "An appraisal of INTERNIST-I," *Artif. Intell. Med.*, vol. 7, no. 2, pp.

93–116, 1995, doi: https://doi.org/10.1016/0933-3657(94)00028-Q.
[11] Sajda Qureshi, "Creating cycles of prosperity with human digital development for intelligent global health," *Inf. Technol. Dev.*, vol. 28, no. 4, pp. 649–659, 2022, doi: https://doi.org/10.1080/02681102.2022.2135872.
[12] S. Ben Zayed, A. Bin Gani, H. F. Gadelrab, and M. K. Bin Othman, "Operational Management in Emergency Healthcare," *Springer* , vol. 297, 2021, doi: 10.1007/978-3-030-53832-3.
[13] S. Ben Zayed, A. Bin Gani, and M. K. Bin Othman, "System Reengineering in Healthcare: Application for Hospital Emergency Departments," vol. 172, 2019, doi: 10.1007/978-3-319-98104-8.
[14] E. R. Rosyid Ridlo Al Hakim, "Android Based Expert System Application for Diagnose COVID-19 Disease: Cases Study of Banyumas Regency," *J. Intell. Comput. Heal. Informatics*, vol. 1, no. 2, 2020, [Online]. Available: https://jurnal.unimus.ac.id/index.php/ICHI/article/view/5958
[15] G. A. Chukwudebe, E. Ekwuwune, and K. I. Nkuma-Udah, "Medical diagnosis expert system for Malaria and related diseases for developing Countries," *2017 IEEE 3rd Int. Conf. Electro-Technology Natl. Dev. NIGERCON 2017*, vol. 2018-January, pp. 24–29, Jul. 2017, doi: 10.1109/NIGERCON.2017.8281875.
[16] J. Munro, "A Review of: 'Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project' Eds Bruce B. Buchanan and Edward H. Shortliffe Addison-Wesley, 1984, 748 pp, £37.00," *Civ. Eng. Syst.*, vol. 1, no. 6, pp. 342–343, Dec. 1984, doi: 10.1080/02630258408970370.
[17] "Artificial Intelligence: A Modern Approach." Accessed: Jun. 17, 2025. [Online]. Available: https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780137505135?srsltid=AfmBOoqb0YgCTklkTw235M61b3Yqugl95hJVcNJLb3R0pgY1e_SIT7fA
[18] Poli Venkata Subba Reddy, "Fuzzy logic based on Belief and Disbelief membership functions," *Fuzzy Inf. Eng.*, vol. 9, no. 4, pp. 405–422, 2017, doi: https://doi.org/10.1016/j.fiae.2017.12.001.
[19] Sukrit Sondhi, ""Mobdoc: Mobile Based Medical Diagnostic Tool,"" Quick Company. Accessed: Nov. 13, 2025. [Online]. Available: https://www.quickcompany.in/patents/mobdoc-mobile-based-medical-diagnostic-tool
[20] Gufran Ahmad Ansari, "An Adoptive Medical Diagnosis System Using Expert System with Applications ," SCISPACE. Accessed: Nov. 13, 2025. [Online]. Available: https://scispace.com/papers/an-adoptive-medical-diagnosis-system-using-expert-system-3tgperdo6b
[21] S. Kumar and A. Kumar, "Automatic Melanoma Detection System (AMDS): A State-of-the-Art Review," *Commun. Comput. Inf. Sci.*, vol. 1214 CCIS, pp. 201–212, 2020, doi: 10.1007/978-981-15-7219-7_17.
[22] P.Santosh Kumar Patra, Dipti Prava Sahu, Indrajit Mandal, "An Expert System for Diagnosis of Human Diseases," *Int. J. Comput. Appl.*, vol. 1, no. 13, 20101, [Online]. Available: https://www.ijcaonline.org/volume1/number13/pxc387439.pdf
[23] J. C. Xinyu Yang, "Clinical use of dendritic cell-derived exosomes for hepatocellular carcinoma immunotherapy: How far we are?," *J. Hepatol.*, vol. 69, no. 4, pp. 984–986, 2018, [Online]. Available: https://www.journal-of-hepatology.eu/article/S0168-8278(18)32183-4/fulltext
[24] C. Uwaoma and G. Mansingh, "Proposing a decision support system for automated mobile asthma monitoring in remote areas*," *Inf. Technol. Dev.*, vol. 24, no. 2, pp. 301–

314, Apr. 2018, doi: 10.1080/02681102.2017.1310712;CSUBTYPE:STRING:SPECIAL;PAGE:STRING:ARTICLE/CHAPTER.

[25] E. J. Roccella, "Report of the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy," *Am. J. Obstet. Gynecol.*, vol. 183, no. 1, 2000, [Online]. Available: https://www.ajog.org/article/S0002-9378(00)40820-3/fulltext

[26] J. Gudu, D. Gichoya, P. Nyongesa, and A. Muumbo, "Development of a Medical Expert System as an ExpertKnowledge Sharing Tool on Diagnosis and Treatment of Hypertension in Pregnancy," *Int. J. Biosci. Biochem. Bioinforma.*, pp. 297–300, 2012, doi: 10.7763/IJBBB.2012.V2.120.

[27] febbyan teguh, "A guide to expert systems," Jan. 01, 1986. Accessed: Nov. 13, 2025. [Online]. Available: https://www.academia.edu/2030055/A_guide_to_expert_systems

[28] E.P.Ephzibah, Dr. V. Sundarapandian, "A NEURO FUZZY EXPERT SYSTEM FOR HEART DISEASE DIAGNOSIS," *Comput. Sci. Eng. An Int. J.*, vol. 2, no. 1, pp. 17–23, 2012, doi: 10.5121/cseij.2012.2103.

[29] N. A. Mahmoud. Alkhayat., "PREVALENCE OF PARENTAL SELF MEDICATION AND ITS POSSIBLE EFFECT TO THEIR CHILDREN IN AL-MADINAH , KINGDOM OF SAUDI ARABIA," *Int. J. Adv. Res.*, vol. 5, no. 1, 2017, [Online]. Available: http://www.journalijar.com/article/14702/prevalence-of-parental-self-medication-and--its--possible-effect-to-their-children-in-al-madinah-,-kingdom-of-saudi-arabia./

[30] X.Y. Djam, Y.H. Kimbi, "Fuzzy Expert System for the Management of Hypertension," *Pacific J. Sci. Technol.*, vol. 12, no. 1, pp. 390–402, 2011, [Online]. Available: https://www.akamai.university/files/theme/AkamaiJournal/PJST12_1_390.pdf

[31] R. O. Guillermo Jorge-Botana, " Call Routing based on a combination of the Construction-Integration model and Latent Semantic Analysis: a full system," Informatica. Accessed: Nov. 13, 2025. [Online]. Available: https://www.researchgate.net/publication/266732933_Call_Routing_based_on_a_combination_of_the_Construction-Integration_model_and_Latent_Semantic_Analysis_a_full_system

[32] Y. Han and R. Ellis, "Implicit knowledge, explicit knowledge and general language proficiency," *Lang. Teach. Res.*, vol. 2, no. 1, pp. 1–23, 1998, doi: 10.1177/136216889800200102;WEBSITE:WEBSITE:SAGE;JOURNAL:JOURNAL:LTRA;ISSUE:ISSUE:DOI.

[33] R. Ellis and C. Roever, "The measurement of implicit and explicit knowledge," *Lang. Learn. J.*, vol. 49, no. 2, pp. 160–175, 2021, doi: 10.1080/09571736.2018.1504229;REQUESTEDJOURNAL:JOURNAL:RLLJ20;SUBPAGE:STRING:ACCESS.

[34] S. Mertens, "Integrated Declarative Process and Decision Discovery of the Emergency Care Process," *Lect. Notes Bus. Inf. Process.*, vol. 409, pp. 85–114, 2020, doi: 10.1007/978-3-030-66193-9_4.