EXPERT SYSTEM USING ARTIFICIAL NEURAL NETWORK FOR CHRONIC RESPIRATORY DISEASES

P. ShubhaDeepti¹, S. V.N. Narayana Rao², V. Naveen Kumar³, Dr. Y. Padma Sai⁴
¹M.Tech Student, ²Associate Professor, ³Assistant Professor, ⁴Professor

Abstract
As per WHO recent estimations, the four major potentially fatal respiratory diseases (Asthma, Pneumonia, Tuberculosis and Chronic Obstructive Pulmonary Disease-COPD) will account for about one in five deaths worldwide. Respiratory diseases are therefore likely to remain a major burden on society for decades to come. Both the prevention and treatment of lung diseases will need to be improved if their impact on longevity and quality of life of individuals, and their economic burden on society, are to be reduced worldwide. Early detection and effective diagnosis is the only rescue to lessen respiratory diseases fatality.

Soft computing approaches are gaining importance in medical disease diagnosis because of their classification performance. In this paper we have developed the expert system which is based on artificial neural network that models human abilities in analyzing and diagnosing respiratory diseases like Asthma, Pneumonia, COPD and Tuberculosis. The objective of developing the system is to diagnosing the potential disease that one may suffer is done by providing relevant inputs through a consultation in which the patient has to answer the set of questions related to signs and symptoms.

Index Terms: Disease diagnosing, respiratory diseases, expert system, inference technique.

I. INTRODUCTION
Respiratory tract diseases are diseases that affect the air passages, including the nasal passages, the bronchi and the lungs. They range from acute infections, such as pneumonia and bronchitis, to chronic conditions such as asthma and chronic obstructive pulmonary disease. Chronic respiratory diseases, including asthma and chronic obstructive pulmonary disease, were responsible for more than 5 million deaths globally [1].

Expert system is a computer program that contains expert knowledge about a particular problem domain, often in the form of if – then rules that is able to solve the problems at a level equivalent to or greater than human expert. Knowledge Engineer collects knowledge from domain expert and transfers it into production rules and creates Knowledge Base. Inference engine then apply different knowledge acquisition techniques and catch the knowledge and deliver it in the form of advice to solve problem. Expert system for diagnosing and recovery monitoring is a specific domain medical knowledge-based system that is used primarily to diagnose respiratory diseases precisely on Asthma, Pneumonia, Tuberculosis and COPD. Based on the diagnosis performed according to the user inputs like symptoms and cough samples, system should be able to provide appropriate disease [2].

The rest of the paper is organized as follows. Respiratory Disease is described in Section II. Architecture of a typical expert system is introduced in Section III. In Section IV, recent research on medical expert system and soft computing technique like Artificial Neural network is discussed. In Section V the design of medical expert systems is presented.
Performance evolution, acknowledgement and conclusion are presented in Section VI and Section VII, respectively.

II. RESPIRATORY DISEASES

Chronic respiratory diseases (CRDs) are of the airways and other structures of the lung diseases like asthma, COPD, tuberculosis, pneumonia etc. They affect more than one billion people worldwide each year and 4 million people die prematurely [3].

Asthma is a common long term inflammatory disease of the airways of the lungs. Symptoms include episodes of wheezing, coughing, chest tightness, and shortness of breath. Sputum may be produced from the lung by coughing but is often hard to bring up. These episodes may occur a few times a day or a few times per week. Depending on the person they may become worse at night or with exercise or cold air. Asthma is thought to be caused by a combination of genetic and environmental factors. Environmental factors include exposure to air pollution and allergens. Other potential triggers include medications such as aspirin and beta blockers. Diagnosis is usually based on the pattern of symptoms, response to therapy over time, and spirometry. Asthma is classified like acute and chronic according to the frequency of symptoms, forced expiratory volume in one second (FEV1), and peak expiratory flow rate [4].

Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium tuberculosis (MTB). Tuberculosis generally affects the lungs, but can also affect other parts of the body. Most infections do not have symptoms, known as latent tuberculosis. About 10% of latent infections progress to active disease which, if left untreated, kills about half of those infected. The classic symptoms of active TB are a chronic cough with blood-containing sputum, fever, night sweats, and weight loss. The historical term "consumption" came about due to the weight loss. Infection of other organs can cause a wide range of symptoms.

Tuberculosis is spread through the air when people who have active TB in their lungs cough, spit, speak, or sneeze. People with latent TB do not spread the disease. Active infection occurs more often in people with HIV/AIDS and in those who smoke. Diagnosis of active TB is based on chest X-rays, as well as microscopic examination and culture of body fluids. Diagnosis of latent TB relies on the tuberculin skin test (TST) or blood tests. Prevention of TB involves screening (screening for disease should begin with a clinical assessment for symptoms suggestive of tuberculosis. Those at high risk include household, workplace, and social contacts of people with active TB. Treatment requires the use of multiple antibiotics over a long period of time [5].

Pneumonia is an inflammatory condition of the lung affecting primarily the microscopic air sacs known as alveoli. Typical signs and symptoms include a varying severity and combination of productive or dry cough, chest pain, fever, and trouble breathing, depending on the underlying cause. Pneumonia is usually caused by infection with viruses or bacteria and less commonly by other microorganisms, certain medications and conditions such as autoimmune diseases. Risk factors include other lung diseases such as cystic fibrosis, COPD, and asthma, diabetes, heart failure, a history of smoking, a poor ability to cough such as following a stroke, or a weak immune system. Diagnosis is often based on the symptoms and physical examination. Chest X-ray, blood tests, and culture of the sputum may help confirm the diagnosis.

Vaccines to prevent certain types of pneumonia are available. Other methods of prevention include handwashing and not smoking. Treatment depends on the underlying cause. Pneumonia believed to be due to bacteria is treated with antibiotics. If the pneumonia is severe, the affected person is generally hospitalized. Oxygen therapy may be used if oxygen levels are low.

Chronic obstructive pulmonary disease (COPD) is a type of obstructive lung disease characterized by long-term poor airflow. The main symptoms include shortness of breath and cough with sputum production. COPD typically worsens over time. Eventually walking upstairs or carrying things will be difficult. Chronic bronchitis and emphysema are older terms used for different types of COPD. The term "chronic bronchitis" is still used to define a productive cough that is present for at least three months each year for two years.
Tobacco smoking is the most common cause of COPD, with a number of other factors such as air pollution and genetics playing a smaller role. In the developing world, one of the common sources of air pollution is poorly vented heating and cooking fires. Long-term exposure to these irritants causes an inflammatory response in the lungs resulting in narrowing of the small airways and breakdown of lung tissue. The diagnosis is based on poor airflow as measured by lung function tests.

Most cases of COPD can be prevented by reducing exposure to risk factors. This includes decreasing rates of smoking and improving indoor and outdoor air quality. COPD treatments include stopping smoking, vaccinations, respiratory rehabilitation, and often inhaled bronchodilators and steroids. Some people may benefit from long-term oxygen therapy or lung transplantation. In those who have periods of acute worsening, increased use of medications and hospitalization may be needed.

Currently to diagnose for pulmonary disorders; X-Ray, computer based tomography (CT) scan, pulmonary tests (routine sputum culture, bronchoscopy, auscultation) are in use, which are very expensive, time consuming and they have serious side effects on human body if they are exposed for longer duration. Also, as part of current diagnoses methods the type of cough or wheeze is just based on observing patient by doctor for a very short period of time. There is no proper system in place to map the appropriate disease based on the patient cough.

**III. EXPERT SYSTEMS**

An expert system is a computer program that emulates the decision making of a human expert, see Figure I.

An expert system is divided into four main categories namely inference engine, knowledge base, working memory and user interface, shown in Fig1. An inference engine is a program that interprets the rules in the knowledge base in order to draw conclusions. A knowledge base is the core strength of an expert system in which the knowledge in the program is expressed in a special-purpose language and kept separate from the code that performs the reasoning. Working memory is a database used to store collection of facts which will later be used by the rules. It is used by the inference engine to get facts and match them against the rules. User interface provides interaction between user of the expert system and the expert system itself. The user of the expert system need not be necessarily an expert in artificial intelligence.

![Figure I: Expert System Components](image)

An inference engine is a computer program that applies artificial intelligence to try to obtain answers or responses to queries from a knowledge base. There are various Techniques used for the development of faster, efficient and reliable medical expert systems like rule based reasoning, decision trees, fuzzy logic, case based reasoning, artificial neural network, genetic algorithm, hybrid approach etc. Following section covers the various types of inference engines [6].

### A. Rule Based Reasoning

This uses "if-then-else" rule statements. The term ‘if’ signifies that the statement is true and the system should ‘then’ carry out a particular function. In case the statement is false or specifically speaking it is ‘not true’, then the ‘else’ statement comes into action and the system has to carry out what the ‘else’ statement tells it to do. In simple rule-based systems, there are two kinds of inference, forward chaining and backward chaining. Rules are simply patterns and an inference engine searches for patterns in the rules that match patterns in the data.

Forward chaining: data gets put into working memory. This triggers rules whose conditions match the new data. These rules then perform
their actions. The actions may add new data to memory, thus triggering more rules. And so on. This is also called data-directed inference, because inference is triggered by the arrival of new data in working memory.

Backward chaining: the system needs to know the value of a piece of data. It searches for rules whose conclusions mention this data. Before it can use the rules, it must test their conditions. This may entail discovering the value of more pieces of data, and so on. This is also called goal-directed inference, or hypothesis driven, because inferences are not performed until the system is made to prove a particular goal (i.e. a question).

B. Decision Trees
A decision tree is a decision support tool that uses a tree-like graph that predicts the classification of a case. The structure includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. A decision tree is a representation which combines search strategy with knowledge relationships. There are various techniques used for developing decision trees like gain ratio, GINI, statistical approaches etc. Decision trees can be evaluated using different techniques. The speed and accuracy of an algorithm can be checked that whether it produces the same results as desired or not. The techniques which can be used to test the performance of decision trees and algorithms are holdout sets, K-fold cross validation, leave 1 out validation, bootstrapping

C. Fuzzy Logic
Fuzzy logic is based on the theory of fuzzy sets, where an object’s membership of a set is gradual rather than just member or not a member. Fuzzy logic uses the whole interval of real numbers between zero (False) and one (True) to develop logic as a basis for rules of inference. Inference methods for fuzzy logic are max-min method, max-dot method, averaged method, root-sum-squared method. Fuzzy decision making procedure consists of three steps, Fig 2:
- Fuzzification: In this step membership degrees for each input variable are computed in respect to its linguistic terms.
- Fuzzy Reasoning: It yields out the output fuzzy set using computed membership degrees and the fuzzy rules. It further consists of rule matching, fuzzy inference, and fuzzy aggregation stages.
- Defuzzification: A crisp value from the output membership function is determined as the final result.

![Figure 2: Fuzzy Inference System](image)

D. Case Based Reasoning
Case-based Reasoning (CBR) is a computational problem solving method of Artificial Intelligence (AI) which works, in contrast to the Knowledge Based System or Rule-Based System, by merely using the successful solutions of previous problems to produce solutions to new and novel problems, without requiring the help of expertise. CBR, like other AI techniques, utilizes a process used by human brain in solving similar problems. If humans face a problem similar to one they have already faced and solved successfully before then they tend to solve the new problem in a similar way too. In the same way, CBR makes use of past cases stored in its case base, to provide solutions to new similar problems. It is like learning from previous experiences. Same has been depicted in Fig 3. Another great advantage of Working by using previous cases is that any new cases which come up, and the solutions of which has to be “revised”, will be saved in the case base for future use, hence automatically updating the case base according to time and changing situations and improving the ability of the CBR system to solve problems as more and more new problems it solves.

E. Artificial neural network
Neural networks are a computational approach which is based on a large collection of neural units loosely modeling the way a biological
brain solves problems with large clusters of biological neurons connected by axons. Each neural unit is connected with many others, and links can be enforcing or inhibitory in their effect on the activation state of connected neural units. Each individual neural unit may have a summation function which combines the values of all its inputs together. There may be a threshold function or limiting function on each connection and on the unit itself such that it must surpass it before it can propagate to other neurons. Typical artificial neural network is displayed in Fig 3.

These systems are self-learning and trained rather than explicitly programmed and excel in areas where the solution or feature detection is difficult to express in a traditional computer program. There are two types of networks like feedforward and feedback.

Neural networks typically consist of multiple layers or a cube design, and the signal path traverses from front to back. Back propagation is the training or learning algorithm where the forward stimulation is used to reset weights on the "front" neural units and this is sometimes done in combination with training where the correct result is known.

![Figure 3: Artificial neural network](image)

**F. Genetic Algorithm**

A genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms. Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection.

At the beginning of the computation a number of individuals (the population) are randomly initialized. The objective function is then evaluated for these individuals. The first/initial generation is produced. If the optimization criteria are not met the creation of a new generation starts. Individuals are selected according to their fitness for the production of offspring. Parents are recombined to produce offspring. All offspring will be mutated with a certain probability. The fitness of the offspring is then computed. The offspring are inserted into the population replacing the parents, producing a new generation. This cycle is performed until the optimization criteria are reached, Fig 4.

![Figure 4: Genetic Algorithm process](image)

- We have used the artificial neural network as it’s widely applied in research because the can model highly non-linear systems in which the relationship among the variables is unknown or very complex.

IV. RELATED WORK

Over the last decade, many interesting techniques of data mining were proposed to detect various types of respiratory diseases. Few of the techniques are described below with their significance and limitations.

Jimmy Singala [7] discussed about the rule based expert system which is based on the SW1-Prolog to diagnose some common...
childhood diseases like asthma, type 1 diabetes, and cystic fibrosis and duchenne muscular dystrophy. The childhood diseases have many common symptoms and some of them are very much alike. This creates many difficulties for the doctor to reach at a right decision or diagnosis. The proposed system can remove these difficulties and it is having knowledge of above mentioned childhood diseases i.e. symptoms are matched to possible disease. Asthma is a childhood disease whose symptoms are wheezing, breathlessness, chest tightness, nighttime’s or early morning coughing. Type 1 diabetes is a disease whose symptoms are increased thirst and frequent urination, extreme hunger, weight loss, fatigue, irritability or unusual behavior, blurred vision and yeast infection. Cystic fibrosis is a disease whose symptoms are salty tasting skin, persistent coughing with and without phlegm, frequent lung infections, wheezing or shortness of breath and poor growth or weight gain. Duchenne Muscular Dystrophy is a disease whose symptoms are delay in walking, frequent falls, and large calf muscles, difficulty in getting up from a lying or sitting position, weakness in lower leg muscles and wadding gait. The user or patient is asked to answer with YES or NO, if a particular symptom appears or not. In the end, based on user’s or patient’s answers, the name of the disease is displayed on the screen. This system is applied on many children and its results are 80% correct.

Krishna Anand [9] has been designed a fuzzy expert system that takes into account details of various patients and identifies the problem the patient is likely to encounter. Besides, the extent of severity of the problem can also be assessed. The expert systems take the following symptoms as input like age, alertness, BMI, cough, difficulty in speaking, fever, nocturnal symptoms, oral steroids, respiratory rate, smoking, time of the day and wheeze. The proper choice of Membership Functions (MF) for each and every parameter plays a pivotal role in determining the efficiency of the system. Triangular MF has been used to represent the parameters like Asthma, Dyspnea, Tuberculosis, Nocturnal Symptoms, Oral Steroids, and Alertness. Trapezoidal MF has been used to represent the parameters like cough, wheeze, difficulty in speaking, respiratory rate, smoking, time of the day). For fever both triangular and trapezoidal MF has been used.

Maryam Zolnoori has developed a fuzzy expert system for prediction of fatal asthma which helps the physicians and patients in order to adopt suitable strategy to prevent the imminent consequences [10]. Fuzzy-rules, modular representation of variables in regard to patients’ perception of the disease, and minimizing the need for laboratory data are the most important features of this system. Main variables are viral infection, exposure to irritants/allergens, duration of asthma, response to treatment, instability of asthma, degree of nocturnal symptoms, degree of intensity of current treatment, peripheral blood eosinophilia, and degree of intensity of exacerbation make antecedents of the rules. Various combinations of inputs and output in general inference networks are formed the production rules of the system. These rules are generated using interview techniques, task performance and protocols, and questionnaires and surveys. Output of this system is possibility of fatal asthma determined in the interval (0-10) i.e. into
two main categories: lower than 5 and more than 5. The category of more than 5 presents patients with the high possibility of fatal asthma that usually include patients with the history of high degree asthma exacerbation. Category of Lower than 5 include patients with low possibility of fatal asthma. These patients are in good conditions regarding the considered variables in this system. To evaluate the results of this system, 10 patients have been interviewed.

V. DESIGN OF EXPERT SYSTEM USING ARTIFICIAL NEURAL NETWORK

A. Data Collection

We have done the enough research to find probable symptoms of the diseases like Asthma, Pneumonia, Tuberculosis and COPD and we have come up with twenty-seven symptoms questioners [11]. Then we visited Sai Siddhartha chest hospital and interacted with patients to collect the symptoms. A total of 60 medical records collected from the patients who are suffering from above mentioned four respiratory diseases, Fig 5.

![Dataset collection from patients in hospital](image)

Figure 5: Dataset collection from patients in hospital

B. Data Organization

Total 27 input variables essential to the diagnosis of the 4 respiratory diseases; these input variables can be divided into three categories: (i) The basic information of a patient like age and the gender (ii) the symptoms, (iii) the history.

In our medical application input nodes represent the symptoms of the four cases of respiratory disease. Such as dry cough, wet cough, fever, wheezing, smoking, weight loss, short of breath, chest pain, dysnea and etc. In addition to this some medical history like latent TB, child hood asthmatic patient, family history of these disease etc. are considered in input data. The output nodes represent four cases of respiratory disease.

Here, each input node corresponds to one input variable, and thus a total of 27 input nodes have been perceived in the input layer, Fig 6. These 27 input variables (categorized into 3 groups) are encoded using the following schemes:

(i) Numerical variables such as age are normalized on to the interval (0, 1). For instance, the patients ages may span from 0 to 100 years old, and thereby the age of a 56-year-old patient can be normalized to the value of 56/100=0.56.

(ii) The variables like family history of disease is with three values, with 0 representing the presence of the Asthma, 1 representing the presence of Tuberculosis, and 0.5 representing the presence of COPD.

(iii) The rest the variables are with two independent attributes, such as the gender, cough, wheezing etc. are encoded with binary values (0, 1). For instance, 1 represents male and 0 female; 1 is adopted when the wheezing is positive, while 0 is used when it is negative.
As we see in our study that the output value is a floating number between 0 and 1, Fig 7.

A. Training, validation and testing the network

Artificial neural networks learn after it faces 60 cases with different symptoms. The computer program was performed under MATLAB software using the neural network toolbox. In the training, the number of neurons on the hidden layer is 10. A dataset including 60 data samples obtained from hospital were used for ANNs. From these, 60% data patterns were used for training the network, 20% for data validation and the remaining 20% patterns were randomly selected and used as the test dataset.

VI. PERFORMANCE EVALUATION

Neural network toolbox from MATLAB 8.5 is used to evaluate the performance of the proposed networks. A two-layer feed-forward network, with sigmoid hidden and softmax output neurons, can classify vectors arbitrarily well, given enough neurons in its hidden layer [12].

The results of applying the artificial neural networks methodology to distinguish between four respiratory diseases patient based upon selected symptoms showed very good abilities of the network to learn the patterns corresponding to symptoms of the person. The training results were very good; the network was able to classify 90% of the cases in the testing set. Same has been shown in Fig.8.

Sensitivity is the proportion of true positives that are correctly identified by a diagnostic test. It shows how good the test is at detecting a disease. Specificity is the proportion of the true negatives correctly identified by a diagnostic test. It suggests how good the test is at identifying normal (negative) condition. Following Table has the overall network sensitivity and specificity values.

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>True Positive / (True Positive + False Negative)</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>True Negative / (False Positive + True Negative)</td>
<td>96.6%</td>
</tr>
</tbody>
</table>

Best validation performance is 0.12012 (mean squared error - MSE) at epoch 10 as shown in Fig.9. The MSE is the average squared difference between outputs and targets. Lower values are better while zero means no error.
VII. ACKNOWLEDGEMENTS

The authors would like to thank doctors of Sai Siddhartha Hospital, Hyderabad, for their support in the respiratory disease symptoms database and valuable suggestions.

VIII. CONCLUSION

Respiratory diseases should be treated well and on time. If they are not treated on time, they can lead to many health problems and these problems may become the cause of death. These problems are becoming worse due to the scarcity of specialists, practitioners and health facilities. In an effort to address such problems, studies made attempts to design and develop expert systems which can provide advice for physicians and patients to facilitate the diagnosis and recommend treatment of patients.

In this paper, we have presented a medical expert system based on the artificial neural network for respiratory disease diagnosis. In particular, we have identified the 27 input variables critical to the diagnosis of the four respiratory diseases like Asthma, Pneumonia, Tuberculosis and COPD and then encoded accordingly. Our Respiratory diseases database consisting of 60 cases has been used in this study. The results show that the proposed system can achieve very high diagnosis accuracy (>90%), proving its usefulness in support of clinic diagnosis decision of respiratory diseases.

REFERENCES