CHAPTER -1
CHAPTER -1

INTRODUCTION

1.1 SOFT COMPUTING

Soft computing (SC) is a mimic human reasoning to make a system intelligent (Akbas et al., 2013). Role model of SC is a human mind. The principal constituents, i.e., tools, techniques of SC are machine learning algorithms (MLA), probabilistic reasoning (PR), fuzzy logic (FL), neural networks (NN) (Sushmita et al., 1996), support vector machines (SVM) (Pan et al., 2014) and evolutionary computation (EC). Soft computing is used to achieve track ability, robustness and low solution cost.

Unique property of soft computing is

- Learning from experimental data.
- These techniques derive their power of generalization from approximating to produce outputs from previous data by using outputs from previous learned inputs.

Applications in real world are

- Application of SC (Paul et al., 2005) to medical diagnoses.
- Application of SC to handwriting recognition.
- Application of SC to image processing & data compression.
- Application of SC to decision-support systems (Ford et al., 1985) and simplify decision trees. (Brewlow and Aha, 1997).

1.2 ARTIFICIAL INTELLIGENCE (AI)

The potential of a tool is to execute functions those are generally linked with a person’s astuteness, such as analysis and optimization all the way through practice. AI is the bough of computer science that attempts to approximate the fallout of human reasoning by
organizing and manipulating truthful and heuristic facts. Areas of AI motion contain expert systems, natural language understanding, speech recognition, vision and robotics.

1.2.1 APPLICATIONS OF AI

- Game playing.
- Speech recognition.
- Understanding natural language.
- Computer vision.
- Expert systems.
- Heuristic classification.

1.3 MEDICAL BACKGROUND OF THYROID DISEASE

The thyroid gland is located on the front part of the neck below the thyroid cartilage (Ambika et al., 2011). The gland produces thyroid hormones, which regulate metabolic rate. Thyroid hormones are important in regulating body energy and body temperature. Diseases of the thyroid gland can result in either production of too much i.e., overactive thyroid or too little which claims underactive thyroid which generates thyroid hormone, thyroid nodules and goiter.

![Thyroid Gland](image)

**Fig 1.1**: Thyroid gland which reasons to Thyroid Disease
Thyroid problems are much more common in women than in children or men. The survey in Section 1.3.1 & 1.3.2 strengthens the statement.

- **THYROID HORMONES PRODUCTION**

  The process of hormone synthesis begins in a part of the brain called the hypothalamus. The hypothalamus releases thyrotropin-releasing hormone (TRH). The TRH travels through the venous plexus located in the pituitary stalk to the pituitary gland, also in the brain. In response, the pituitary gland then releases thyroid-stimulating hormone (TSH) into the blood. The TSH travels to the thyroid gland and stimulates the thyroid to produce the two thyroid hormones, L-thyroxine (T4) and triiodothyronine (T3). The thyroid gland also needs adequate amounts of dietary iodine to produce T4 and T3, the molecules of which contain four and three atoms of iodine, respectively.

- **THYROID HORMONES REDUCTION**

  To prevent the overproduction or underproduction of thyroid hormones, the pituitary gland senses how much hormone is in the blood and adjusts the production of hormones accordingly. For example, when there is too much thyroid hormone in the blood, TRH and TSH productions are both decreased. The sum effect of this is to decrease the amount of TSH released from the pituitary gland and to reduce production of thyroid hormones from the thyroid gland to restore the amount of thyroid hormone in the blood to normal. Defects in these regulatory pathways rarely may result in hypothyroidism or hyperthyroidism. The most common cause of hypothyroidism and hyperthyroidism occurs due to problems within the thyroid and not the regulatory system.

- **THYROID GOITER**

  Thyroid goiter is an enlargement of the thyroid that can occur with hyperthyroidism or hypothyroidism but also with benign and malignant nodules. Worldwide, the most common cause of goiter is iodine deficiency. It is now less common with the use of iodized salt. Multiple nodules in the thyroid are common, but only about 5% of the nodules are a thyroid cancer. Thyroid cancer rates have been increasing steadily by about 6% every year for more
than 20 years. It is one of the few cancers whose rate is increasing and whose very low rate of mortality is also rising with time.

1.3.1 THYROID DISEASE EFFECT

When the thyroid produces too much hormone, the body uses energy faster than it should. This condition is called hyperthyroidism. When the thyroid doesn't produce enough hormones, the body uses energy slower than it should. This condition is called hypothyroidism. There are many different reasons why either of these conditions might develop. Currently, about 20 million residents of West have some form of thyroid disease. People of all ages and races can get thyroid disease. However, women are five to eight times more likely than men to have thyroid problems.

In India too, there is a significant burden of thyroid diseases. According to a projection from various studies on thyroid disease, it has been estimated that about 42 million people in India suffer from thyroid diseases.

The Indian Council of Medical Research established the National Cancer Registry Program (NCRP), and the NCRP has collected the data of more than 3, 00,000 patients between the periods 1984 and 1993. Among these patients, the NCRP noted 5614 cases of thyroid, and this included 3617 females and rest 2007 child & males. The seven centers involved in the studies were at Mumbai, Delhi, Thiruvananthapuram, Dibrugarh, Chandigarh, Andhra and Chennai. Among them, Andhra & Thiruvananthapuram had the highest relative frequency of cases of thyroid among all cases enrolled in the hospital registry, 1.99%, 1.97% among males and 5.71%, 5.76% among females respectively. The nationwide relative frequency of thyroid cancer among all the cancer cases was 0.1%-0.2%. The age-adjusted incidence rates of thyroid cancer per 100,000 are about 1 for males and 1.8 for females as per the Mumbai Cancer Registry, which covered a population of 9.81 million records. The histological types of thyroid cancer were studied in a Hospital Cancer Registry of 1185 "new cases" of thyroid.

1.3.2 THYROID FUNCTION IN PREGNANCY AND CHILDREN

In the past 2 years, exciting work has been carried out to understand the thyroid in pregnancy and childhood in India. The normal reference ranges of TSH, T4U and FTI in pregnant women are
used the 5\textsuperscript{th} and 85\textsuperscript{th} percentile to define the reference ranges in the disease-free records. The reports of the trimester-wise values in the first, second and third trimesters were TSH(3) (1.92-5.86, 3.2-5.73 and 3.3-5.18), T4U(4) (12-19.45, 9.48-19.58 and 11.32-17.7) and FTI (0.6-5.0, 0.44-5.78, and 0.74-5.7), respectively. This is with regard to iodine deficiency during a normal or an abnormal creep of initial or at extreme months of period in pregnancy due to the hormonal growth.

In a large epidemiological study of 24,685 students from all over India published from 2010, the define normative data on thyroid hormone levels in healthy children is calculated as the reference ranges from 5343 to 8346. After excluding children who had a personal or family history of thyroid disease, used thyroid medications, had a goiter and had hypoechogenicity or nodularity on ultrasound or positivity for serum anti-TPO antibodies. The authors reported that for TSH, the 97th percentile was in the range 6.01-8.4 for boys and 5.28-8.04 for girls. This, as the authors themselves suggest, offers a compelling argument against lowering the reference range of TSH in this population.

The above study clearly ensures and gives an idea, thyroid effect levels are more in women and at next level to children and at last level, it is with men.

1.3.3 MAJOR CAUSES OF THYROID DISEASE

There are several different causes of thyroid disease.

The following conditions cause hypothyroidism:

- Thyroiditis is an inflammation of the thyroid gland. This can lower the amount of hormones produced.
- Postpartum thyroiditis occurs in 5\% to 9\% of women after giving birth. It is usually a temporary condition.
- Iodine deficiency is a problem affecting approximately 100 million people around the world. Iodine is used by the thyroid to produce hormones. Although prevalent before the 1950s, iodine deficiency has been virtually wiped out by the use of iodized salt.
A non-functioning thyroid gland affects one in 4,000 newborns. If the problem isn't corrected, the child will be physically and mentally retarded. All newborns are given a screening blood test in the hospital to evaluate thyroid function.

**The following conditions cause hyperthyroidism**

- With Graves' disease, the entire thyroid gland might be overactive and produce too much hormone. This problem is also called diffuse toxic goiter.
- Nodules might be overactive within the thyroid. A single nodule is called toxic autonomously functioning thyroid nodule, while several nodules are called a toxic multinodular goiter.
- Thyroiditis, a disorder that can be painful or painless, can also release hormones that were stored in the thyroid gland causing hyperthyroidism for a few weeks or months. The painless variety occurs most frequently in women after childbirth.
- Excessive iodine is found in a number of drugs such as Amiodarone, Lugol's solution (iodine) and some cough syrups might cause the thyroid to produce either too much or too little hormone in some individuals.

**1.3.4 CURRENT DIAGNOSES PRACTICES**

Thyroid disease can be difficult to diagnose because symptoms are easily confused with other conditions. Fortunately, there is a test called the thyroid stimulating hormone (TSH) test, that can identify thyroid disorders even before the onset of symptoms.

The Journal of the American Medical Association found that screening for mild thyroid failure in women and men over age 35 is as cost-effective as screening for more common problems such as high cholesterol or high blood pressure.

When thyroid disease is caught early, treatment can control the disorder even before the onset of symptoms. The goal of treatment for any thyroid disorder is to restore normal blood levels of thyroid hormone.
Hypothyroidism is treated with a drug called levothyroxine. This is a synthetic hormone tablet that replaces missing thyroid hormone in the body. With careful monitoring, the doctor will adjust the dosage accordingly, and thus return to normal lifestyle.

Hyperthyroidism, generally more difficult to treat, requires the normalization of thyroid hormone production. Treatment could involve drug therapy to block hormone production, radioactive iodine treatment that disables the thyroid, or even thyroid surgery to remove part or the entire gland.

The most popular treatment is radioactive iodine. This therapy often results in hypothyroidism, requiring the use of levothyroxine in order to restore normality.

Thyroid diseases are life-long conditions with careful management, people with thyroid disease can also live healthy, normal lives preventing it at earlier stages. The detailed information about the datasets and the attribute values are reported in Chapter 3.

1.4 EXPERT SYSTEM

Expert System (ES) in general simulates both the knowledge and the knowns of human experts. i.e., the ES solves troubles that are usually solved by human experts. All expert systems include at least three basic rudiments; a knowledge base (Babu et al., 2010), information about the subject, an inference engine where the data is to be known and altered, and a user interface is the ability of creating a human share his inputs through the system and the system constructs the output which is necessary.
**Fig 1.2: Expert system elements architecture.**

The elements of the expert system are shown in **Fig 1.2**

**Expert:** Expert is a familiar person who has the information and practice about particular domain.

**Knowledge Base:** Knowledge Base (KB) is used to store complex data.

**Inference Engine:** The inference engines were components of expert system which create adaptation ability to change or alter the data.

**User interface:** The user interface (UI) designed in an information device with which a human being may interact, visualize and input the data from elsewhere it is given.

**User:** Users normally use a system without the technical expertise required to fully understand it.

**1.5 RULE BASED SYSTEM**

**Rule based System (RBS)**

These are knowledge representation centers *(Aqdas and Qamar, 2015)* on the use of IF condition THEN action statement.
If the situation matches the **IF** part of a rule, the action specified by the **THEN** part of the rule is performed. This action may affect the outside world. Direct programs control or may instruct the system to reach a conclusion. This matching of rules to the facts can produce inference chains.

**Inference engine**

The inference engine is the software structure that locates knowledge and infers new knowledge from the obtainable knowledge. Many different paradigms ([Guangquan et al., 2012](#)) are used in expert system, but we used one of the two primary concepts.

**Backward chaining**

It is a top-down reasoning process that starts from the desired goals and works backward requisite conditions.

**Forward reasoning**

It is a bottom up reasoning ([Babu et al., 2011](#)) process that starts with known mechanism towards the desired goal.

**Users**

There are three different user modes for an expert system

- Getting answers to problems.
- Improving or increasing the system’s knowledge.
- Sharing the KB for human users.

**User Interface**

The job of the interface is to exchange information between the operator and the inference engine. The ability to trace responsibility for conclusion by providing expert data for problem solving. The explanation subsystem can trace such responsibility.
1.6 MACHINE LEARNING ALGORITHMS

Machine Learning Algorithms (MLA) is a technical restraint that is anxious with the plan and growth of algorithms \textit{(Aha et al., 1991)} that let computers to be trained based on data. The chief spotlight of machine learning investigation is to robotically learn the method or the ability to distinguish multipart patterns and make intellectual decisions based on data. Hence, machine learning is directly related to the fields of computer science.

1.6.1 MOTIVATION TOWARDS MACHINE LEARNING

- Machine learning is a subfield of computer science where there is a study based on facts.
- Accuracy levels are elevated in prophecy, when the data is reliable.
- Deals with riotous data (Ambiguity and Errors).
- It has a strong tie with AI & Optimization.
- It examines dataset.
- Analyzing the data is named as Simplification & Classification.

Machine learning is composed of two components namely

(a) Training Phase (Labeled data)
Use some algorithms and encode training examples which are obtained from previous experiences.
(b) Prediction Phase (Classification through unlabeled data)
This is used to predict that of the class which is a new instance obtained, not exactly relevant trained data.

\textbf{Fig 1.3: Trained and test datasets.}

11
1.6.2 APPLICATIONS

Machine learning includes:

- Machine perception, computer vision, natural language processing.
- Syntactic pattern recognition and search engines.
- Medical diagnosis, bioinformatics and brain-machine interfaces.
- Game playing, software engineering and robot locomotion.

1.6.3 CLASSIFICATION OF ALGORITHMS

Data organization is a two-step process namely; Model Construction and Model Usage. In model construction, a classifier is built narrating a predetermined set of data classes or concepts (Erdem et al., 2012). In learning phase a categorization algorithm builds the classifier by analyzing or a guidance set made up of database and their linked class labels. In model usage phase, the performance of classifier model is evaluated with the known label of test samples for classifying a new tuple.

The classification techniques which does not belongs to the above two categories are

1.6.3.1 The classification methods

- Bayesian classification.
- Decision tree classification.
- Rule-based classification.
- Classification by back propagation.
- Support vector machines.
- Classification based on association rule mining.

Lazy learner’s classification methods are

- K-nearest-neighbor.
- Case based reasoning.

The other classification methods are

- Genetic algorithms.
1. Rough sets.
2. Fuzzy logic techniques.

1.7 RESEARCH CONTEXT AND SCOPE OF THE THESIS

The research context and scope of the thesis is presented in the form of deliverables and boundaries given below.

1.7.1. DELIVERABLES

1. The performance of evolutionary algorithms and machine learning algorithms are presented in this thesis.
2. A novel approach, modified similarity relation using RST had made the accuracy and the performance of prediction high for evolutionary algorithms.
3. Accurate results from consistent data through MLA for result analysis using new proposed approach decisive factors.
4. Comparative study of statistical significant difference applied on TDA using ANOVA method.
5. Thyroid Disease Diagnosis expert advisory system (TDDEAS) presented for providing an expert advice.

1.7.2. BOUNDARIES

1. The classification and evolutionary algorithms on RST presented in this thesis for accurate diagnosis of TD are suitable to thyroid patient data only.
2. Regularization methods presented in this thesis works effectively for the TDA based on the feature sets or attributes.

1.8 PROBLEM STATEMENT

This research work is aimed at TDD by using rough sets and MLA. An attempt has been made to develop an ES namely TDDEAS to identify the thyroid disorders and to suggest necessary precautions.

THE MAIN OBJECTIVES OF THE PRESENT THESIS ARE GIVEN BELOW

- To study the performance of evolutionary algorithms based on RST for better diagnose.
Examine the possibility of developing a novel technique.

- To explore the important features from TDA for further improving the performance of selected classifiers.
- To investigate the opportunities of identifying the best classifier in MLA, a new approach of decisive factors invoked.
- A comparative study on UCITD and IETD data to explore the problems in datasets.
- To prepare an expert advisory system for TDD.

1.9 ORGANIZATION OF THE THESIS

The thesis is divided into seven chapters as outlined below:

Chapter 1: This chapter discusses the brief introduction to basic concepts along with its Research context & Scope, Proposed Methodology, Objectives and Organization of thesis. In this chapter the basic idea towards the primary methods are displayed.

Chapter 2: This chapter deals with a comprehensive review of the literature along with a survey on Thyroid Disease Diagnose, rough set theory, Machine Learning Algorithms and Evolutionary Algorithms.

Chapter 3: This chapter deals with collection of thyroid data from UCI machine learning repository and Indian eTDML repository. This chapter gives an idea regarding rules applied on the KB. This chapter in advance deals with different categories and pristine attributes of thyroid disease.

Chapter 4: In this chapter, TDDEAS is generated by considering the primary concepts of SMS, RBS and later evolutionary algorithms ABC and PSO are applied separately to check the levels of classification for predicting thyroid disease, further the rough sets theory is used to find the missing attribute value which evolves in a novel approach of modified similarity relation, which is useful for further scope of prediction and classification of the disease and suggestion of referral source.

Chapter 5: In this chapter, MLA is introduced into the concept for ensuring that the output obtained is of high accuracy. This chapter includes the implementation of methods like regularization methods, which has ridge regression algorithm, least absolute shrinkage and
selection operator and proposed new method decisive factors.

**Chapter 6:** In this chapter, the comparative study for two different datasets UCITD and IETD are produced, and the study shows the significant values obtained between groups and within groups through ANAVO statistical method, which gives the scope to this thesis work and a justification to computer science in building a decision making system for classification and prediction based on separate datasets.

**Chapter 7:** The overall conclusion towards the work is explained in this chapter along with its future scope. The conclusion part clearly mentions, how useful each and every attribute in classification and prediction of TD. This chapter is followed by appendix, references and authors publications list.

**1.10 SUMMARY**

This chapter gives a beginning to the basic fundamentals which inspired choosing the machine learning platform to work on the modules. Each and every concept which is explained in this **Chapter 1** is thoroughly used in developing and deploying the system which is useful for human kind. The problem statement in this chapter clearly gives the information, the procedure of making the proposed system. The **Section 1.9** clearly represents the actual work which is proposed in this thesis.