International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Computational way of Classifying Thyroid Disorder Patients and Identifying the Similar Patients using a Novel Method

S. Jamuna, K. Mohan Kumar

Abstract*: The thyroid hormones secreted by thyroid gland are interrelated with many metabolic processes of our body. Any dysfunction of thyroid gland leads to thyroid diseases. Hypothyroidism and hyperthyroidism are the very common thyroid disorders which affect the large number of people nowadays. Prediction of thyroid diseases at right time and giving suitable medicines to the patients help them to overcome the health problems. A machine learning technique will definitely assist the physicians for the prediction and treatment of thyroid diseases. In this work, the datasets are taken from UCI repository and Fuzzy- C Means algorithm is used for the clustering the thyroid diseases.

Keywords: Cluster analysis, clustering, Thyroid, Fuzzy C-Means, Hypothyroidism, Hyperthyroidism and Prediction.

I. INTRODUCTION

The butterfly shaped thyroid gland which is located at low and front part of our throat below the Adam's apple secrets some vital hormones such as Triiodothyronine(T3), Thyroxin(T4) collectively known as thyroid hormones play an active role for proper functioning of our body including normal metabolism, body temperature, heart rate, growth and development, etc. Adequate secretion of thyroid hormones is essential for infants and children for the development of their brain. Thyroid stimulating Hormone(TSH) secreted in pituitary gland regulates the secretion of thyroid hormones. Abnormal secretion of thyroid hormones causes many health problems in our body and they are called as thyroid disorders that may be from a small goitre to dangerous cancer [1].

The major two thyroid disorders which commonly affect the human beings especially women are hyperthyroidism and hypothyroidism. Overproduction of thyroid hormones leads to hyperthyroidism where as underproduction results in hypothyroidism. As thyroid hormones make impact on each and every cell of our body, thyroid disorders severely affect the normal functioning of our body depending upon their level [2, 3].

Revised Manuscript Received on March 30, 2020.
***** Correspondence Author
S.Jamuna*, Research Scholar, PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.
Dr. K. Mohan Kumar, Head, PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

So diagnosis of thyroid disorders at early stage is very needful one for better treatment. Data mining is one of the most successful processes for identifying, classifying and grouping huge datasets into desired information. There are so many data mining techniques like association rules, classification, clustering, etc used by healthcare organizations for the extraction of information from the pre-processed or raw data. Fuzzy c- means is one of the best algorithms for clustering that accepts a particular range of data to two or more clusters and gives the accurate result [4].

The three objectives of this paper are clustering similar category of patients based on the types of thyroid disorders, finding the similar patients and study the percentage of patients affected by various kinds of thyroid disorders in the dataset. These works are done with the help of a developed software tool.

II. LITERATURE REVIEW

By using data mining techniques different kinds of works have been done for thyroid disorders. Ankita Tyagi et al.(2018) used Machine learning algorithm, Support Vector Machine(SVM), KNN and Decision Tree with the dataset containing minimum number of parameters taken from UCI machine learning repository for the prediction of estimated risk of a person's chance to get thyroid disease[5].

Using EM clustering algorithm and J48 classification algorithm A.Sumathi et al.(2018) created a tool not only for differentiating the thyroid diseases but its subtypes also. They availed the additional attributes RT3 and Basel metabolic temperature for the diagnosis of the subtypes of hypothyroidism [6].

S.Sathya Priya, and Dr.D.Anitha et al.(2017) analysed and compared four classification models namely Naïve Bayes, Decision Tree, Multilayer Perception and Radial Basis Function Network and found that Decision Tree model provided the best rate of classification to identify the common thyroid dysfunction[7].

Vikram V Hegde and Deepamala N (2016) created a computerized prediction system which was developed by using Artificial Neural Networks(ANN) to classify the thyroid dysfunctions as hypothyroid, hyperthyroid and normal with the accuracy rate of 85%[8].

Dr.G.Rasitha Banu et al.(2015) utilized DBSCAN3[Density Based Spatial Clustering of Applications with Noise] to predict the thyroid disease with related symptoms. DBSCAN assumed all the parameters in the dataset as unassigned then it classified them using Hierarchical multiple classifier classification scheme to give accurate result to the users[9].

Shivanee Pandey et al. (2013) used various data mining techniques such as Bayes net, Multilayer perception, RBF network, C4.5, CART, REP tree, Decision Stump to diagnose hypothyroid disease using WEKA software with accuracy of 99.6%[10].

III. METHODOLOGY

The thyroid dataset is collected from UCI repositories. This dataset deals with 215 patients. The following Table I shows the various parameters which are related to thyroid disorders.

Table I: Thyroid dataset parameters

S.No	Parameters
1.	Patient's ID
2.	Gender
3.	Age
4.	T3 Uptake Test
5.	Thyroid Stimulating Hormone (TSH)
6.	Thyroxine(T4)
7.	Triiodothyronine(T3)
8.	TSH Value After Injection

The following Table II shows the datasets collected from UCI repository.

Table II: Dataset

PATINT ID	GENDER	AGE	T3 UPTAKE TEST	TRIIODOTHYRONINE(T3)	THYROXINE(T4)	THYROID STIMULATING HORMONE(TSH)	TSH VALUE AFTER INJECTION(200 mg thyrotrophic)
1	Male	20	107	2.02	10.1	0.9	2.7
2	Female	25	113	2.01	9.9	2	5.9
3	Female	35	127	1.05	12.9	1.4	0.6
4	Female	24	109	0.8	5.3	1.4	1.5
5	Male	23	105	0.9	7.3	1.5	-0.1
6	Female	45	105	2.01	6.1	1.4	7
7	Male	65	110	1.94	10.4	1.6	2.7
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
214	Female	54	97	0.54	4.7	8.7	12.6
215	Male	52	102	0.45	4.5	6.4	6.7

The following Table 3 shows the normal thyroid test condition [11].

Table III: Normal Thyroid levels

TEST	Abbreviation	Typical Ranges
Triiodothyronine	**T3**	**0.7 - 2.04 ng/dl**
Thyroxine	**T4**	**5.1 - 14.1 µg/dl**
Thyroid Stimulating Hormon	**TSH**	**0.270 - 5.35 µIU/ml**

Where ng- nanogram, µg-microgram, µIU- micro International Unit The dataset in Table II classified into three groups such as normal, hyperthyroidism and hypothyroidism using the following Table 4.

Table IV: Type of Thyroid [12]

TSH	T4	T3	Result
High	Normal	Normal	Normal
Low	Normal	Normal	Normal
High	Low	Low	Hypothyroidism
Low	High	High	Hyperthyroidism

A. KARL PEARSON CORRELATION

It is a statistical method used to find the similarity between linear related objects [13, 14]. In this work this formula is used to find the similarity between the new patient and the existing set of patients. The correlation r is calculated by the following formula.

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Where

Σdx =Sum of deviations of X series from its Assumed Mean *i.e.* $\Sigma (X-A_x)$

Σdy =Sum of deviations of Y series from its assumed Mean *i.e.* $\Sigma (Y-A_y)$

Σdx^2 =Sum of squared deviations of X Series from its Assumed Mean *i.e.* $\Sigma (X-A_x)^2$

Σdy^2 =Sum of squared deviations of Y Series from its AssumedMean *i.e.* $\Sigma (Y-A_y)^2$

$\Sigma dxdy$ =Sum of products of deviations of X and Y series from their respective assumed means. $\Sigma dxdy = \Sigma (X-A_x)(Y-A_y)$

N =Number of pairs

B. PROPOSED MODEL

The proposed model has been implemented using PHP and MYSQL codes and run in Microsoft Windows environment. The parameters Triiodothyronine (T3), Thyroxine(T4) and Thyroid Stimulating Hormone (TSH) play an important role to determine the thyroid disorders. So, in this tool it is used to diagnose and classify the thyroid persons. The thyroid framework model shown in the following Fig. 1, finds the three stages of thyroid.

The main objective of the proposed framework is finding the stage of thyroid disorders using T3, T4 and TSH values and classifying the patients based on all the parameters in the thyroid dataset.

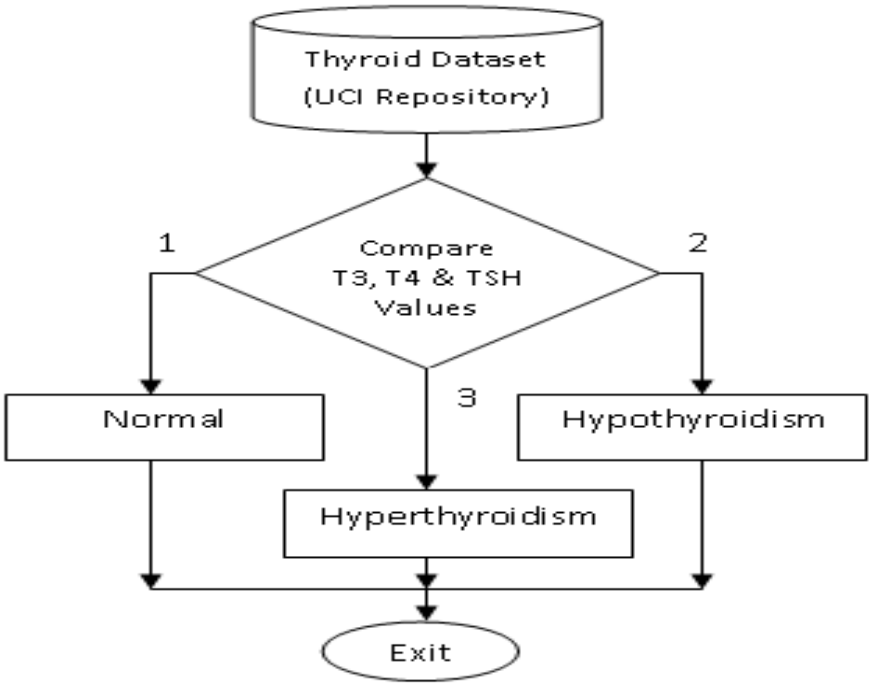

Fig. 1. Thyroid Classification Model

The following Fig. 2 shows the similarity prediction for the new patient in thyroid framework model. This similarity is done by applying Karl Pearson correlation formula which is used to help the medical practitioner to prescribe the medicines.

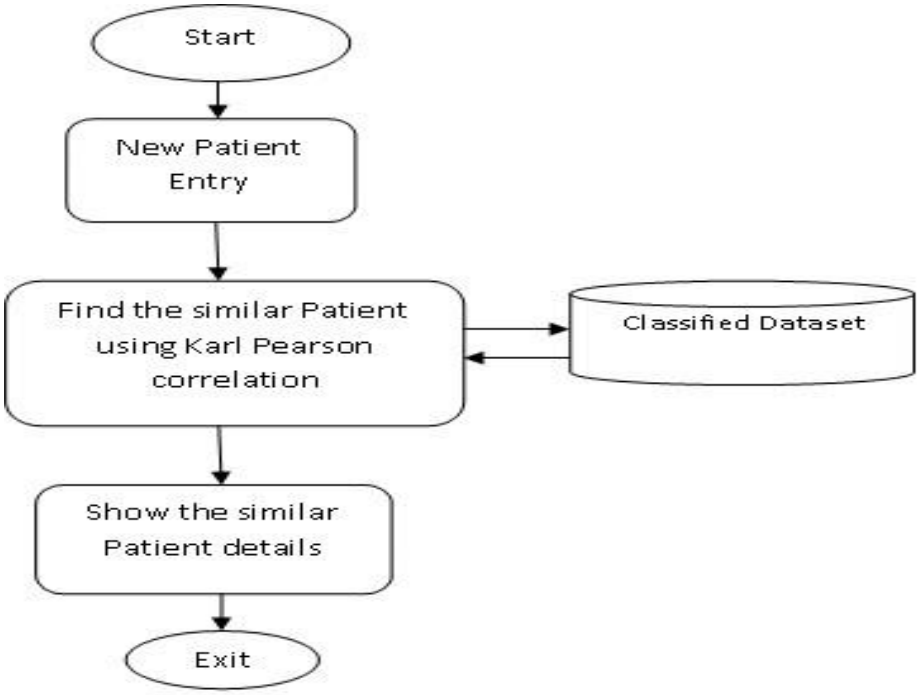

Fig. 2. Similarity Prediction Model

IV. RESULT AND DISCUSSION

The following Fig. 3 shows the screen shot of the login window of the developed software tool.

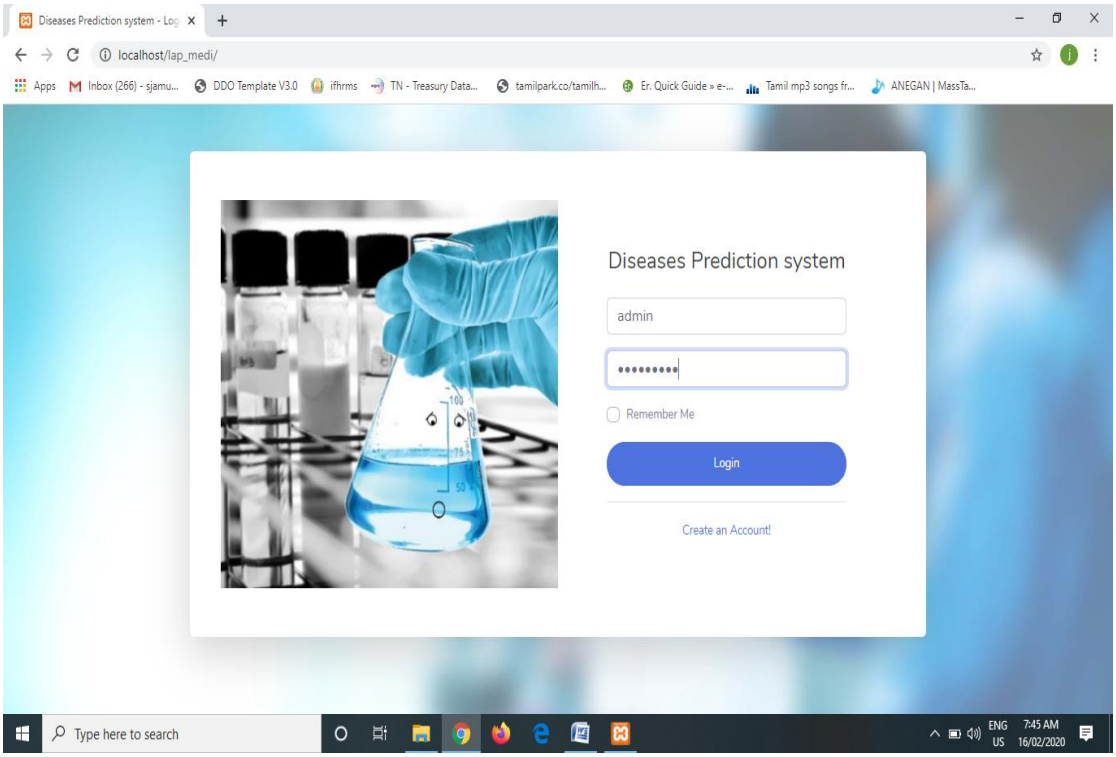

Fig. 3. Login window

The following window shown in Fig. 4 is used to upload the collected dataset from UCI repository for thyroid.

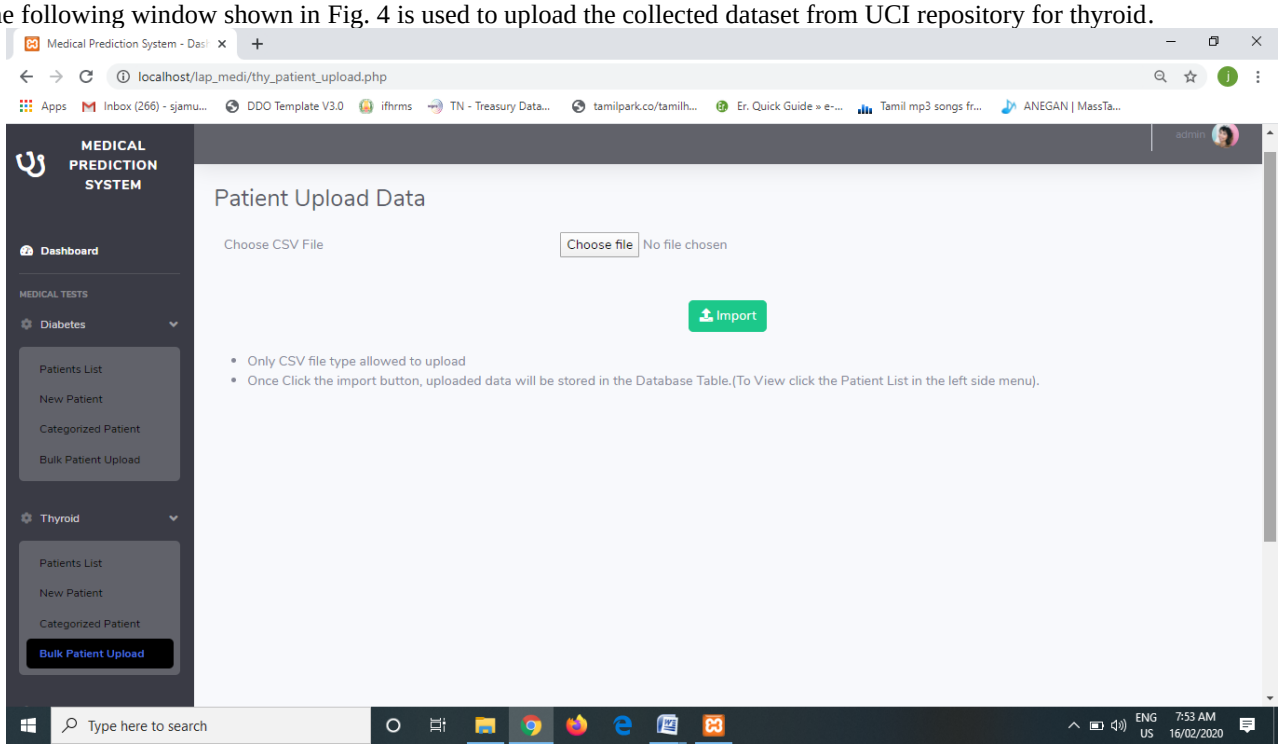

Fig. 4. Data upload option window

The following Fig. 5 shows the complete dataset which are uploaded.

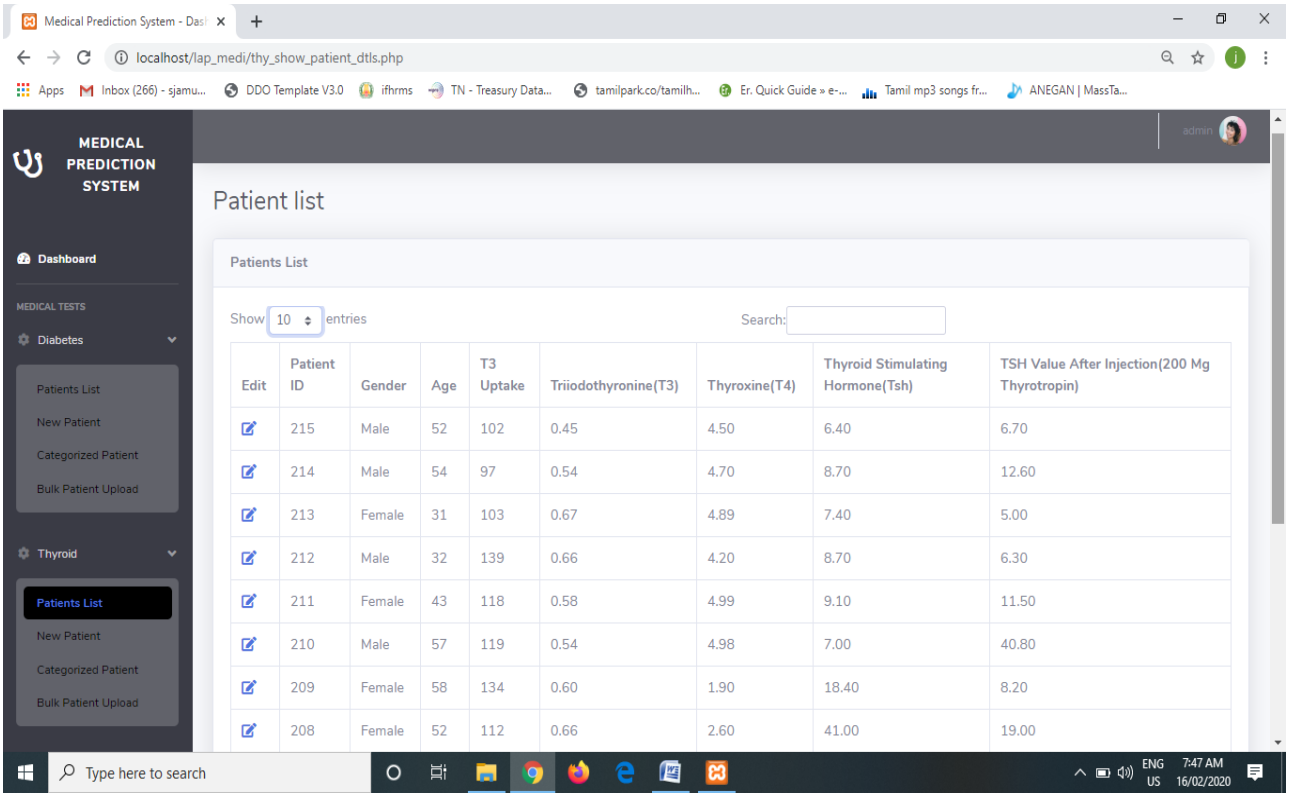

Fig. 5. Uploaded dataset window

The following window in Fig. 6 shows the classification of patients based on the three categories.

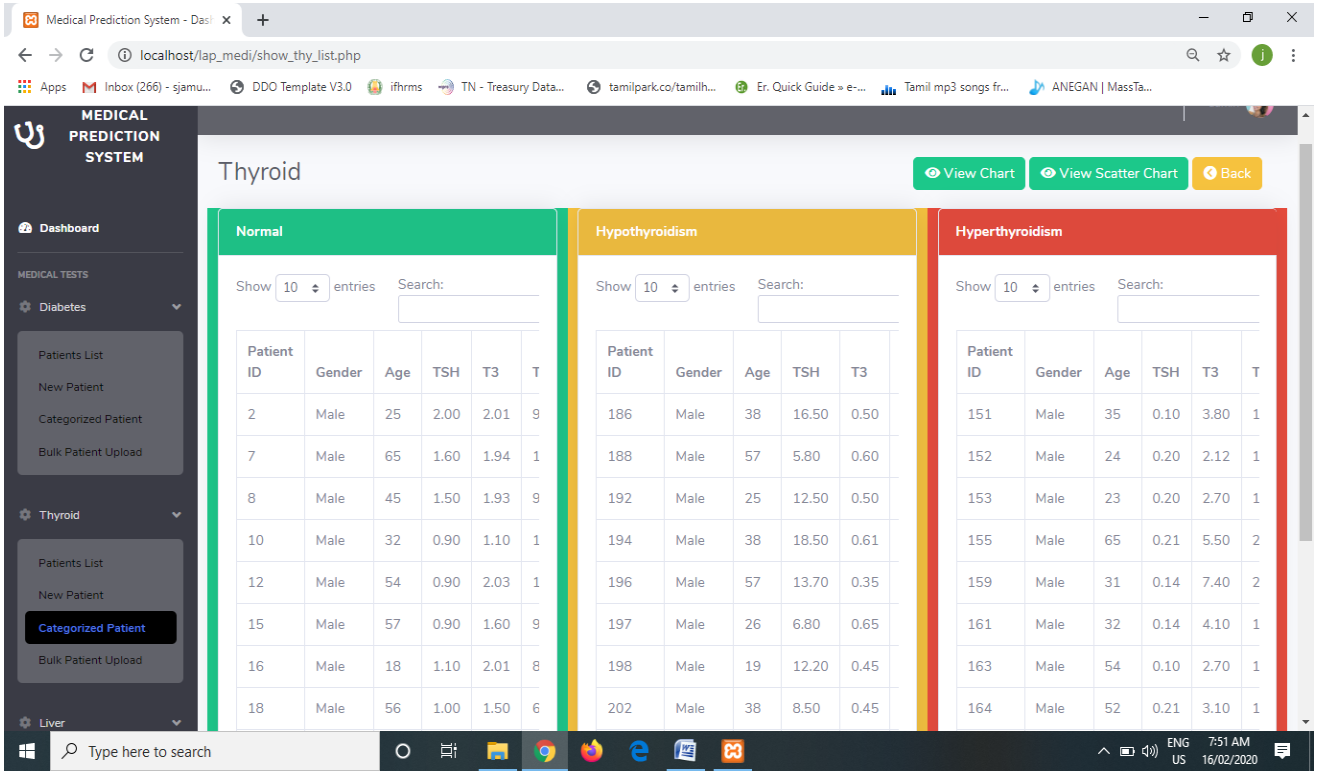

Fig. 6. Classification of patients

The following Fig. 7 shows the window which is used to upload a new patient's information .

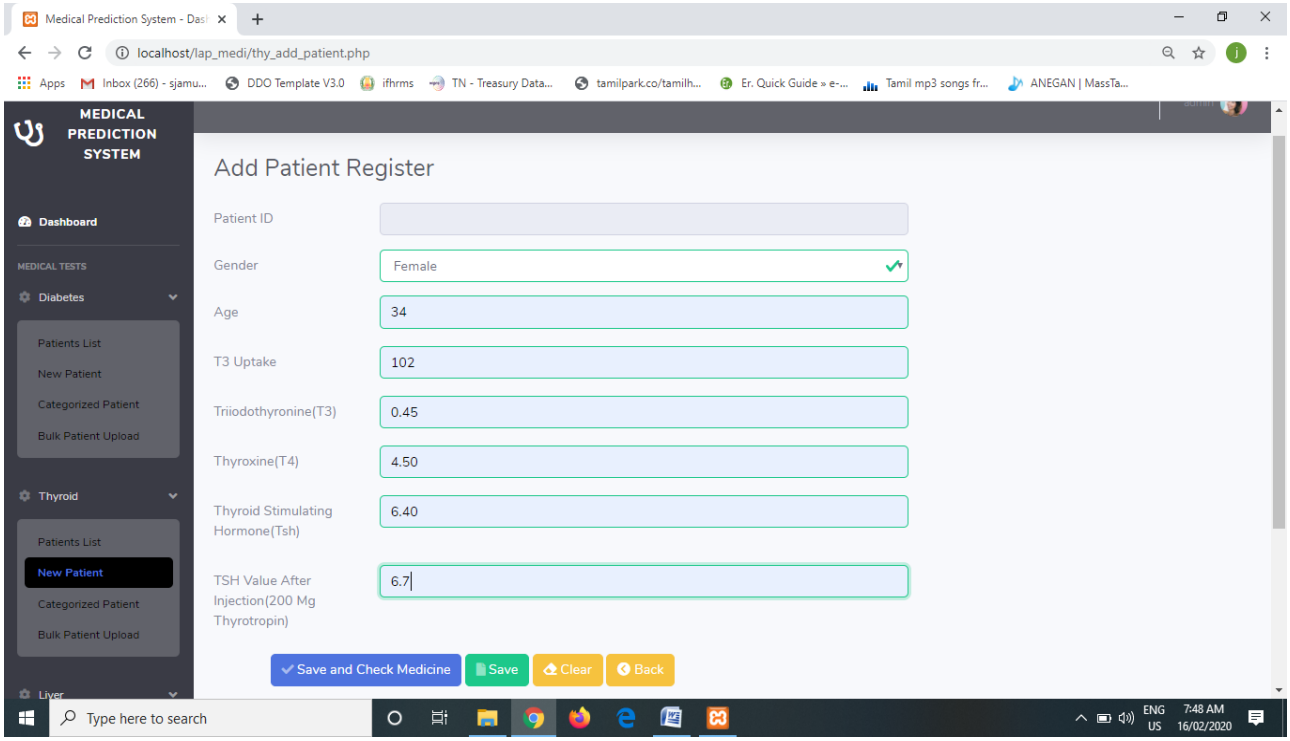

Fig. 7. Upload a new patient's information

When the medical practitioner enters a new patient, this tool finds the similar existing patient using Karl Pearson correlation formula. The following Fig. 8 gives the similar patient details. The doctor can prescribe the same medicine to the new patient using the suggestion given by the tool.

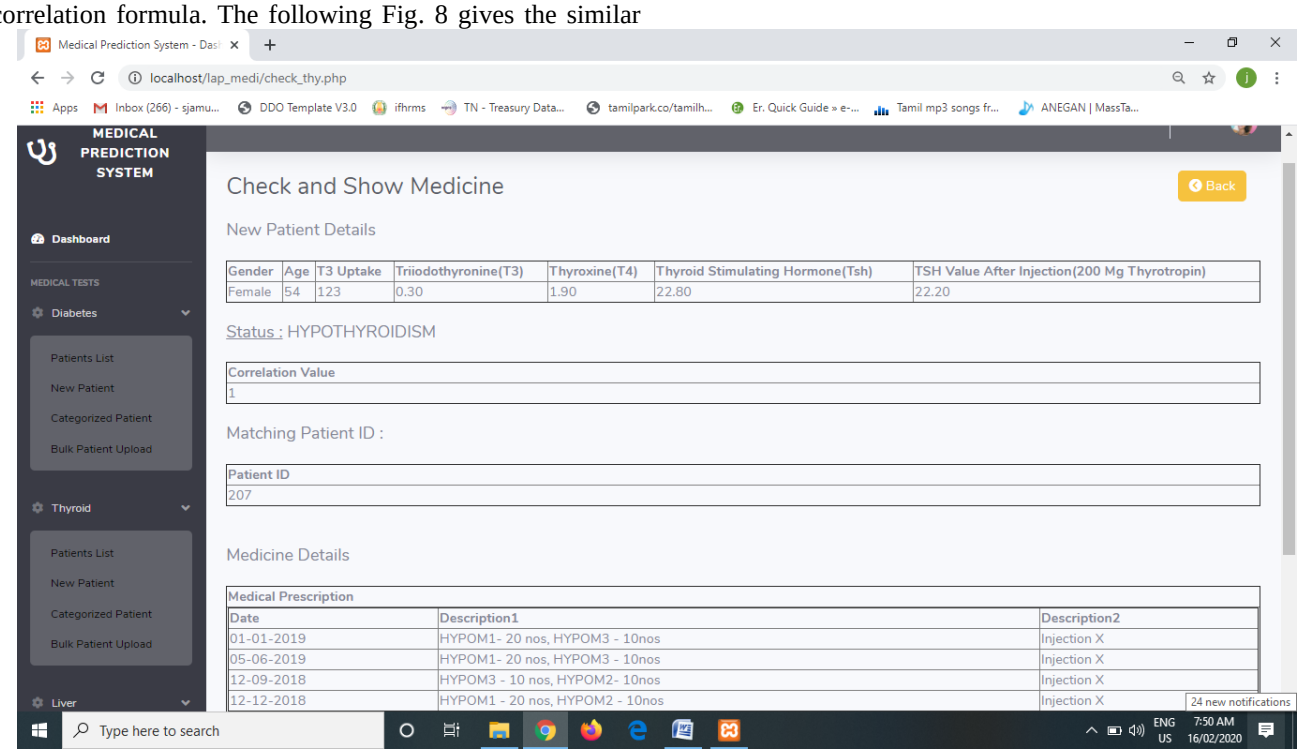

Fig. 8. Similar patient's detail

The following Fig. 9 shows the Overall thyroid patients' percentage of the given dataset. In this clustering algorithm FCM is used [15,16]

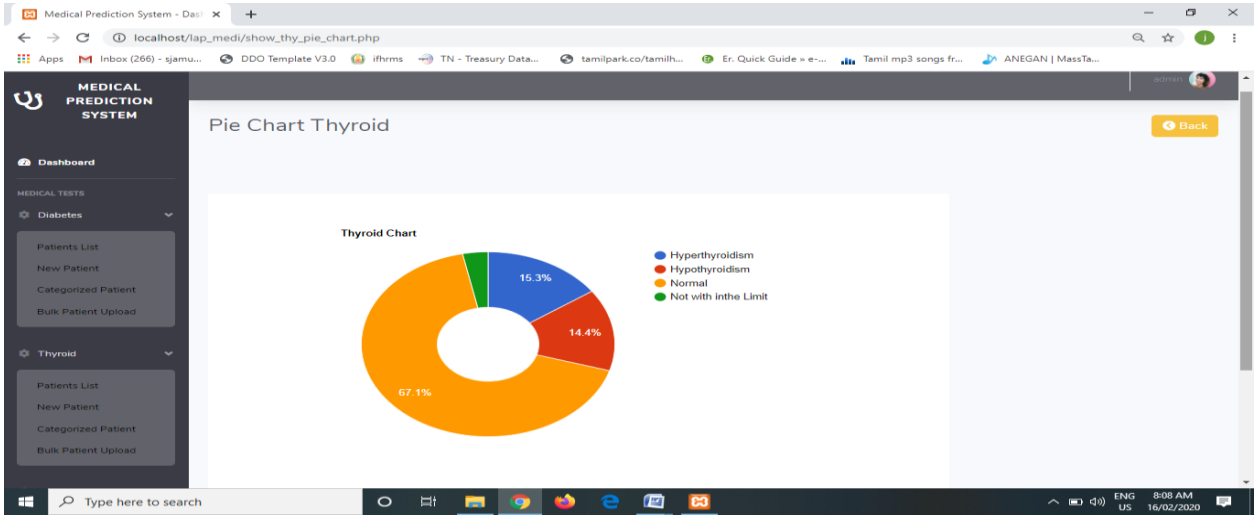

Fig. 9. Overall thyroid patients' percentage

The above Fig. 9 clearly shows 67.4% of people are in normal level, 14% of people in hypothyroidism, 15.3% of people are in hyperthyroidism. Here 3.3% of data is not relevant for this analysis.

V. CONCLUSION

This paper vividly shows the prediction and classification of thyroid disorders using FCM algorithm and Karl Pearson correlation with high accuracy rate for the dataset given as input. This will help the physicians, research scholars, medical students and attendants for the analysis of thyroid disorders. The tool will be enhanced in future for other diseases also like anaemia.

REFERENCES

1. Https://www.healthline.com/health/ hypothyroidism/ symptoms-treatments-more.
2. Http://www.emedicinehealth.com/thyroidfaqs/article-em.htm.
3. Http://www.foxnews.com/health/2012/ 2/10/hypothyroidism-versus-hypert hyroidism.html (accessed dec 2015).
4. Deepika Koundle,et al.,"Computer- Aided Diagnosis of Thyroid Nodule:A Review", International Journal of Computer Science and EngineerinSurvey(IJCSES) Vol.3,No.4,Aug 2012.
5. Ankita Tyagi, Ritikamehra and Aditya Saxena, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique", 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018), 20-22 Dec, 2018, Solan, India.
6. Sumathi A, Nithya G and Meganathan S, "Classification of Thyroid Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 13881-13890 ISSN: 1314-3395 (on-line version) url: http://www.ijpam.eu Special Issue.
7. S. Sathya Priya and Dr. D. Anitha, "Survey on Thyroid Diagnosis using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Special Issue 1, January 2017,ISSN (Online) 2278-1021 ISSN (Print) 2319-5940.
8. Vikram V Hegde and Deepamala N,"Automated Prediction of Thyroid Disease using ANN ", International Journal of Innovative Research in Science, Engineering and Technology,(An ISO 3297: 2007 Certified Organization) Vol. 5, Special Issue 10, May 2016.
9. Dr. G. Rasitha Banu, M.Baviya, and Dr.Murtaza Ali, "A Study On Thyroid Disease Using Datamining Algorithm", International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 3, Issue4 -July-August 2015, PP. 376-379.
10. Shivanee Pandey,Rohit Miri, S. R. Tandan, "Diagnosis And Classification Of Hypothyroid Disease Using Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,Vol. 2 Issue 6, June – 2013.
11. Carl A. Burtis, Edward R.Ashwood, aDavid E.Bruns, and Barbara G.Sawyer, "Tietz, Fundamentals of Clinical Chemistry", 6th Edition Hardcover – Abridged 2008.
12. Http://Mp-Body/Articles-1 / Everything-you- want- to- know – about-the-thyroid.
13. Http://www.economicsdiscussion.net /correlation/karl-pearsons-formula- for-finding-the-degree-of-correlation/2579.
14. Jennifer D.Chee, "Pearson's Product Moment Correlation: Sample Analysis" University of Hawaii at Mānoa School of Nursing on May 29, 2015.
15. Dr. K. Mohan Kumar and S. Jamuna, "Comparative study on Machine learning Clustering Algorithms using Medical related Datasets", International Journal of Management, Technology And Engineering, Volume IX, Issue III, MARCH / 2019.
16. K. Mohan Kumar and S. Jamuna, "Prediction of Diabetes and Clustering Based on Its Levels Using Fuzzy C Means Algorithm", International Journal of Scientific & Technology Research, VOL 9, ISSUE 2, February 2020 Edition.

AUTHORS PROFILE

S.Jamuna, pursed M.Sc & M.Phil degree in Computer Science from Bharathidasan University, Tiruchirappalli, Tamil Nadu,India. Currently doing Ph.D as a part time research scholar in PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur, Affiliated to Bharathidasan University, T.N, India. She is having 13 years of teaching experience. Her main research work focuses on Data Mining.

Dr. K. Mohan Kumar received M.Sc & Ph.D degree in Computer Science from Bharathidasan University, Tiruchirappalli and M.Phil degree in computer science from Manonmaniam Sundaranar University, Thirunelveli, India. At present works as head in PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur, T.N, India. His main research work focuses on Network Security, Machine Learning, IoT and Data Mining, published more than 50 research papers in reputed International journals. He has 25 years of teaching experience and 20 years of Research Experience.