Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 12 August 2022

Prediction Of Thyroid Disease Using Support Vector Machine Approach

Dr.P.A.Selvaraj, Dr.M.Jagadeeshan, A. NithinKumar

Department Of Computer Applications

Kongu Engineering,Perundurai Erode-638060,Tamil Nadu,India

ABSTRACT:

Thyroid gland is anendocrine gland and one of the human body's most significant organs. This gland secretes two hormones that aid in the regulation of the body's metabolism.

Hyperthyroidism and Hypothyroidism are the two forms of thyroid diseases. When this illness occurs in the body, specific hormones are released, causing the body's metabolism to become unbalanced. Thyroid issues A blood test is used to identify this condition, however the results are frequently unclear and noisy. To make the data rudimentary sufficient for the details to reveal the patient's riskcontracting the condition, data purification procedures were applied. In illness prediction, machine learning is extremely important.

Using (SVM)Support Vector Machine, this study demonstrated the intuition of how to forecast thyroid illness and emphasise the important qualities for the ailment. The thyroid data set from the machine learning repository was utilised for this research.

Keyword : Support Vector Machine, Thyroid Disease, Min Max Scalar

Introduction:

Thyroid illness affects at least ten persons out of every 100 in India. Thyroid illness affects a large number of women between the ages of 17 and 54. Different phases of a thyroid disease in its severe state include cardiac difficulties, raised blood pressure, elevated cholesterol levels, lethargy, and negative effects on fertility.

Total serum thyroxin (T4) and total serum triiodothyronine (T3), which maintain the body's metabolism, are produced by the thyroid gland.The hormones are necessary for the normal functioning of every cell, organ, and tissue, as well as for general energy generation and control, protein synthesis, and body temperature regulation.

The concept of thyroid sickness diagnosis and therapy is represented by the inner working behaviour of the

thyroid disease, which is a critical point in many thyroid illnesses.Depending on whether hormone levels are normal, excessive, or defective, thyroid disease is classified as euthyroidism,hyperthyroidism,or hypothyroidism.

Curing disease is a constant concern for health care practitioners, and getting the right diagnosis for the patient at the right time is crucial.

Thanks to many better diagnosis technologies, the old medical report is becoming outdated and is being replaced with a report based on symptoms.Computational biology has evolved and is now employed in the healthcare business. There are algorithms that can be used to diagnose the disease at an early stage. There are many datasets in medical information systems, but only a few intelligent systems that can readily analyse illness.Over time, machine learning algorithms have become increasingly important in fixing these complicated and indeterminate challenges in the growing model.Any illness forecast models are used to overrule the attributes that may be picked from various datasets and employed in the most accurate categorization of healthy patients. Misclassification can result in a healthy patient receiving unneeded therapy if this is not done. As a result, we conclude that the Support Vector Machine technique makes text processing on data fairly simple. It processes massive amounts of data quickly and with the highest level of accuracy possible.

DEFINITION OF THE PROBLEM:

Thyroid diseases are on the rise in India, according to data. Thyroid problems affect about one out of every ten Indian adults. Thyroid illness affects an estimated 42 million individuals in the United States.

Predicting thyroid disorders by a doctor is a time-consuming procedure that might result in a poor outcome; only an experienced doctor can fully assess the situation. To aid medical surgeons and support them in prediction of illness and reduce their workload by machine learning algorithms.

OBJECTIVE:

- The primary goal is to create a system that can anticipate which form of thyroid illness a patient is suffering from.
- To forecast thyroid illness with the fewest possible parameters.

LITERATURE SURVEY:

[3]They used Logistics Regression and SVM machine learning techniques to investigate the Thyroid Dataset. A comparison was made between these two approaches based on Precision, Recall, F measure, ROC, and RMS error.The best classifier turned out to be Logistic Regression.

[2]Solan, India, December 20-22, 2018. International Meet in Parallel, Distributed, and Grid Computing Using Machine Learning Techniques.They employ a variety of categorization techniques in this study, including other algorithms likeSVM, Artificial Neural Networks, and the *knn*algorithm. Classification and detection were done based upon the data set got from the UCI Repository the accuracy will be determined according to the output produced. By comparing the algorithmspart of accuracy employed to identify the optimal approach with the highest accuracy.

[5]To forecast cancers, a Neural Network of feed forward is employed toUltrasound feature extraction and segmentation pictures. All of the average scores for exactness and any other variables that are more than 86 percent.

[4]Thyroid nodules are detected as benign or cancerous utilising ultrasound pictures of the thyroid and image processing techniques such as radiomics and deep learning. Between these two techniques, a comparison is conducted. The radiomics-based method's classification accuracy, sensitivity, and specificity are 66.81 percent, 51.19 percent, and 75.77 percent, respectively, whereas the deep learning-based method's evaluation indexes are 72.61 percent, 62.12 percent, and 81.21percent, correspondingly. Machine learning was shown to be the most effective method.

SYSTEM ANALYSIS AND DESIGN:

METHODOLOGY:

The thyroid data set for the first section is collected from an open source repository.Thehypothyroidism dataset is utilised. Before feeding these data sets to training, they must be verified.Null data or extraneous data may be present such data are being eliminated by data cleaning.Data that has been cleaned is utilised as both training and test data for the algorithm. The method pulls characteristics from various datasets in order to categorise data according to labels. Test data is given into the system to ensure that the forecast is accurate. By comparing the characteristics of both, probability will be calculated for test data based on the feature retrieved.Regardless of whether the label is hypothyroidism or not, the greatest probability value will be applied to it.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 12 August 2022

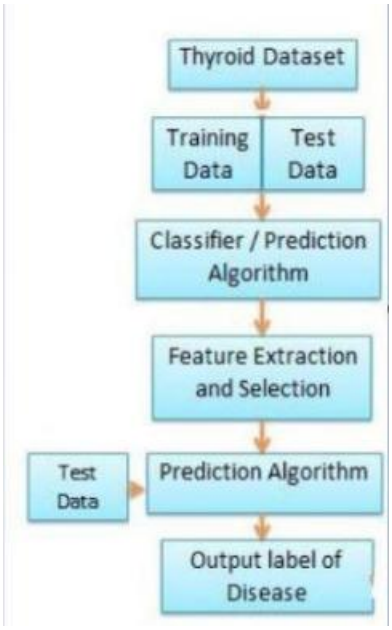


Fig 1:Work Flow Chart

DATASET:

A fresh thyroid dataset with 3772 samples and 14 characteristics is now available in an open source machine learning repository.

```
thyroid_df = pd.read_csv('../input/thyroid-disease-data-set/hypothyroid.csv')
thyroid_df.head()
```

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	query hypothyroid	...	TT4 measured	TT4	T4 me
0	41	F	f	f	f	f	f	f	f	f	...	t	125	t
1	23	F	f	f	f	f	f	f	f	f	...	t	102	f
2	46	M	f	f	f	f	f	f	f	f	...	t	109	t
3	70	F	t	f	f	f	f	f	f	f	...	t	175	f
4	70	F	f	f	f	f	f	f	f	f	...	t	61	t

Fig 2:Dataset Header

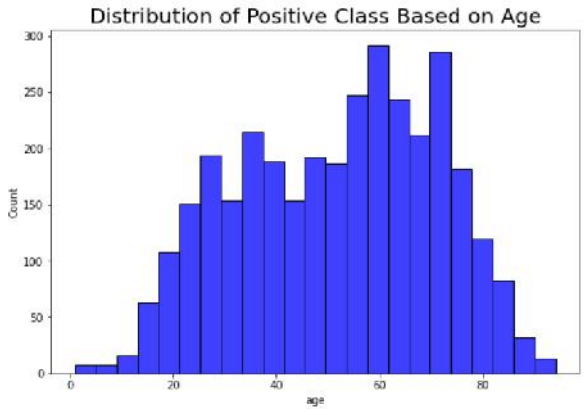
Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 12 August 2022

EVALUATION AND TESTING:

HISTOGRAM CHART:

```
plt.figure(figsize=(9,6))
sns.histplot(x='age',data=positive_df,color='blue')
plt.title("Distribution of Positive Class Based on Age",{'fontsize':20});
```



The distribution of a numeric variable for one or more groups is represented by a histogram. The data is divided into bins, with each bin represented by a bar. This website contains a collection of histograms created in Python using the seaborn and matplotlib tools. This histogram depicts the age range of persons who are afflicted by thyroid illness.

COUNTPLOT CHART:

```
sns.countplot(x='Label',data=thyroid_df)
plt.title("Countplot for Target variable");
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 12 August 2022

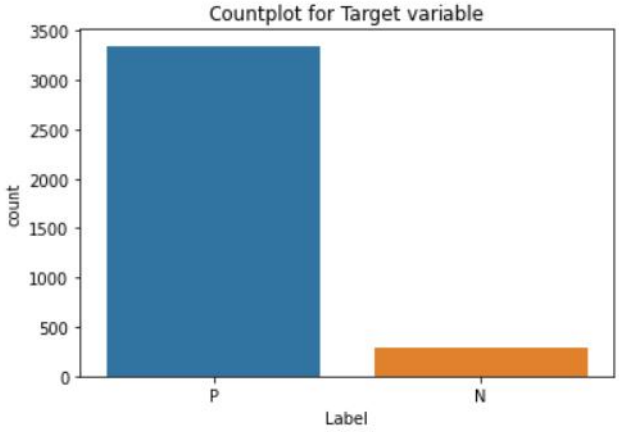


Fig 4:Countplot Chart

A count plot is a histogram across a category variable rather than a quantitative variable. You may compare counts across nested variables using the same basic API and settings as for barplot(). The number of positive and negative examples for the target variable are displayed in a count plot chart.

PIE CHART:

```
plt.figure(figsize=(10,8))
plt.pie(x=positive_df.sex.value_counts(),
        labels=['Female','Male'],
        startangle = 90,
        colors=['springgreen','orange'],
        autopct='%.2f'
       );
plt.legend();
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 12 August 2022

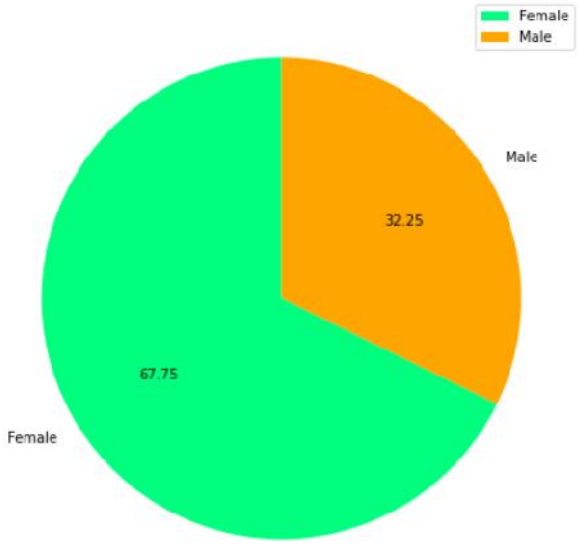
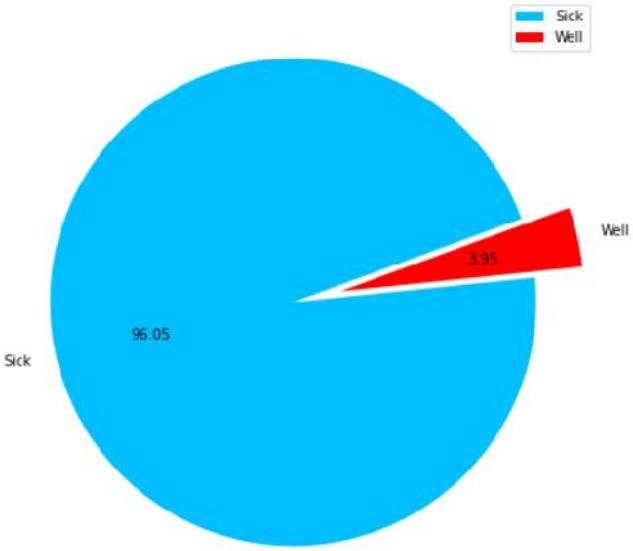


Fig 5:Percentage of Male and Female

A pie chart is a spherical statistical layout that can only show one set of data. The overall percentage of the provided data is represented by the chart's area. The proportion of sections of the data is represented by the area of the pie slices. Pie wedges are the pieces of the pie. The percentage of men and women with thyroid disease symptoms is depicted in this graph.



This pie chart represents the percentage number of people sick and well from thyroid.

Fig 6:Percentage of Sick and Well

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 12 August 2022

HANDLE IMBALANCE DATA:

```
plt.figure(figsize=(8,8))
plt.pie(x=positive_df.sick.value_counts(),
        labels=['Sick','Well'],
        startangle = 20,
        colors=['deepskyblue','red'],
        autopct='%.2f',
        explode=[0,0.2]
       );
plt.legend();
```

```
smote = SMOTE(random_state=11)

x_smote, y_smote = smote.fit_resample(X_train, y_train)
```

```
print("Shape before the Oversampling : ",X_train.shape)
print("Shape after the Oversampling : ",x_smote.shape)
```

```
Shape before the Oversampling :  (2896, 14)
Shape after the Oversampling :  (5340, 14)
```

Fig 7: Handling Imbalance Data

SMOTE is one of the most extensively utilised oversampling techniques for resolving the imbalance problem synthetic minority oversampling technique.Its purpose is to produce a moreequitable distribution of classes by randomly reproducing examples of minority classes. SMOTE creates new minority instances by combining

MINMAX SCALAR:

```
scalr = MinMaxScaler()
x_smote.TT4 = scalr.fit_transform(x_smote[['TT4']])
x_smote.age = scalr.fit_transform(x_smote[['age']])
x_smote.FTI = scalr.fit_transform(x_smote[['FTI']])
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 12 August 2022

The MinMaxscaler scales the minimum and highest values to 0 and 1, respectively. The StandardScaler, on the other hand, scales all values between min and max to fit inside a range of min to max. First, a MinMaxScaler object with default hyperparameters is created. Once defined, we can use the fit transform() method to build a transformed version of our dataset by passing it our dataset.

Support Vector Machine(SVM):

This is a well-known machine learning algorithm which is used to solve issues in both classification and regression. The purpose of SVM is to find a hyperplane in an X-dimensional space (X being the number of features) that best separates the data points. The 'Hinge loss' feature assists in margin maximisation.

```
models = {
    SVC():"Support Vector Machine"
}
for m in models.keys():
    m.fit(x_smote,y_smote)
for model,name in models.items():
     print(f"Accuracy Score for {name} is : ",model.score(X_test,y_test)*100,"%")
```

```
Accuracy Score for Support Vector Machine  is :  98.20441988950276 %
```

CONCLUSION:

Using Machine Learning the thyroid detection that focus to develop a smart and exact method of identifying thyroid disease. To train our dataset and more accurately identify thyroid illness, we used the (SVM) Support Vector Machine technique. Based on the user's input, the computer is taught to recognise whether the person is normal or has hypothyroidism. Our goal is to provide community with an efficient and exact method to machine learning approach that might employed towards illness detection applications.

FUTURE ENHANCEMENT:

Further research might be done by applyingprocessing of high ultrasonic image inspection of thyroid pictures byanticipatingthe nodulesand cancer that aren't visible in blood tests.
Thyroid disease prediction can encompass all thyroid-related disorders by integrating both data.

REFERENCES:

1. Chaubey, Gyanendra, et al. "Thyroid disease prediction using machine learning approaches." *National Academy Science Letters* 44.3 (2021): 233-238.
2. Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." *Fifth international conference on parallel, distributed and grid computing (PDGC)*.
3. Vasan, Chandan R. Chetan, et al. "THYROID DETECTION USING MACHINE LEARNING." *Computing (PDGC-2018)*.
4. Wang, Yongfeng, et al. "Comparison study of radiomics and deep learning-based methods for thyroid nodules classification using ultrasound images." *Ieee Access* 8 (2020).
5. Garg, Hitesh, and Alka Jindal. "Segmentation of thyroid gland in ultrasound image using neural network." *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*.
6. Raisinghani, Sagar, et al. "Thyroid prediction using machine learning techniques." *International Conference on Advances in Computing and Data Sciences*. Springer, Singapore, 2019.