



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 4, April 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.488**

 9940 572 462

 6381 907 438

 [ijircce@gmail.com](mailto:ijircce@gmail.com)

 [www.ijircce.com](http://www.ijircce.com)

# Enhanced and Detection of Human Disease Symptoms Prediction using Data Mining Categorization

P.Charumathi,<sup>1</sup> Mrs. K.K.Kavitha,<sup>2</sup> Mrs. D.Latha,<sup>3</sup>

Research Scholar, Department of Computer Science, Selvamm Arts and Science College (Autonomous),  
TamilNadu, India<sup>1</sup>

Head of the Department, Department of Computer Science, Selvamm Arts and Science College (Autonomous),  
TamilNadu, India<sup>2</sup>

Assistant Professor, Department of Computer Science, Selvamm Arts and Science College (Autonomous),  
TamilNadu, India<sup>3</sup>

**ABSTRACT:** "Human scale, type, structure, and category of imaging" are powerful components in computational biology, particularly for detection and segmentation. The ultrasonic method provides the most accurate prognosis on which kind of thyroid nodule is present. Images were separated into two classes, in this study: those that were fully distinguishable from those that were not. Benign (non-cancerous) and Malignant (cancerous) Thyroid Nodules. I found these objects were helpful in classifying the sample data. To devise an automated classification system that utilises machine learning to classify the nodules and segments the nodule area. There are several classifiers, including the SVM, the KNN, and the Bayesian classifier. There are 13 extracted features which are assigned to classification algorithms like SVM, KNN, and GLCM. The finding corresponds with radiologic photographs from its source and evaluation, such as correctness was calculated. It is shown that the SVM has superior results to Bayesian classifiers.

**KEYWORDS:** Thyroid and Cancer Disease, Support Vector Machine, Decision Tree, Naïve Bayes.CNN Algorithm, Classification. Meridian Filter.

## I.INTRODUCTION

As of now private and open medical services associations are giving enormous measures of information which are dreary to deal with. Because of effective scientific strategies for distinguishing obscure and significant data, information mining has turned out to be prominent in healthcare domain. It has been utilized by the medical practitioners to distinguish and cluster for deriving patterns among extensive number of factors, thus enabling in disease prediction using recorded databases. Data mining is the most fundamental step of research with the goal of obtaining important data from enormous data indexes.

These data mining techniques results in efficient medical diagnosis and optimal medical data management. These techniques also help in analyzing the factors that are responsible for the cause of specific diseases like habits, food, lifestyle and so on. Worldwide the most typical type of cancer women suffer is Breast cancer. The malignant tumour grow in breast tissue cells, which undergoes partition and develop abnormally causing cell division and cell death, hereby cancer on breast tissue is called breast cancer. Universally it is one of the most well known kind of disease that is influencing approximately 10% of all ladies at some phase of their life. Despite the fact that scientists couldn't identify the appropriate reason for the disease, they were able to predict certain risk factor that aggregates the probability of breast cancer development in women. Some of the elements are age, family history risk et cetra.

With early diagnosis, upto 97% of women survive about five more years. Heart disease is a major health issue for a long time. The term heart disease alludes to substantial number of coronary illness. They analyse the abnormal health conditions that specifically influences the functions of heart. Coronary illness comprises a lot of

conditions that influences the heart, some of the ailments of heart are irregular heart beat or heart valve variations from norm.

Diabetes is one of the persistent medical illnesses. Almost in every age group diabetes disease is common. The blood glucose level gets elevated, due to deformity in insulin secretion. Insulin is the hormone that pays way for glucose to arrive and unlocks the cells of the body. The diabetic person metabolism does not produce sufficient insulin. So a diabetic person is inclined to danger of infected by alternate diseases such as nervous breakdown, blood vessel damage, renal failure and loss of vision. Its effect is growing rapidly and its treatment is expensive making commoner to suffer. Globally most typical endocrine disease is Thyroid disease. Thyroid disorders are dissimilarities in stipulated regular functioning of the thyroid gland thereby causing irregular hormone secretion resulting in hyperthyroidism and hypothyroidism. Medical analysts report that their symptoms are similar to other disorders. The functional dataset of thyroid gland is relatively important for accurate diagnosis and prediction of the diseases associated with the gland. They help in regulating body metabolism.

## II.LITERATURE SURVEY

**2.1A Deep-Learning Algorithm for Thyroid Malignancy Prediction from Whole Slide Cytopathology Images**, David Dov the propose over view a deep-learning-based algorithm that is inspired by the way cytopathologist diagnoses the slides.(1,2-4) TBS 2 indicates a benign slide, TBS 3, 4and 5 reflect inconclusive findings with an increasing risk of malignancy, and TBS 6 indicates malignancy. TBS 1 is assigned to inadequately prepared slides and is out of the scope of this work. (1, 45, 46). **Thyroid Diagnosis from SPECT Images Using Convolution Neural Network with Optimization**, Laying Ma. Convolution neural network (CNN) is a typical depth learning model, which consists of a series of modules. ) research of this paper, we improve the network architecture and the training method based on Dense Net model. ) experimental results show that this method has higher accuracy and better performance than other models in thyroid disease classification using SPECT image.(2,40-50). In the [SVM and PNN were used to classify hypothyroidism and hyperthyroidism. Feature selection is done by using GA. The comparative study was done on two datasets from DCI. Both PNN and SVM gave same efficient accuracy result of 100% on dataset one, but for second dataset classification accuracy of SVM was much better 99.02 % than PNN 96.8 %.( 3, 14-16). **USING ARTIFICIAL NEURAL NETWORK IN DIAGNOSIS OF THYROID**, Farad Soleimanian Gharehchopogh. In this paper, we consider a Multi-layer Perception (MLP) ANN using back propagation learning algorithm to classify Thyroid disease. It consists of an input layer with 5 neurons, a hidden layer with 6 neurons and an output layer with just 1 neuron(4,2-3).**Prediction of Thyroid Disease Using Data Mining Techniques**, Irina IoniNă. four classification models: Naive Bayes, Decision Tree, Multilayer Perception and Radial Basis Function Network. The results indicate a significant accuracy for all the classification models mentioned above, the best classification rate being that of the Decision Tree model. The data set used to build and to validate the classifier was provided by UCI machine learning repository and by a website with Romanian data. The framework for building and testing the classification models was KNIME Analytics Platform and Weka, two data mining software.(5,6-9).

## III.EXISTING SYSTEM

Thyroid is one of the largest endocrine gland. It is a small butterfly shaped gland which is located in the front portion of the neck. It is located just below the Adams apple. Thyroid gland produces hormones that help the body to control metabolism. There are various thyroid disorders. It includes Hyperthyroidism, Hypothyroidism, goiter and thyroid nodules (benign/malignant). Various modalities that are used to detect and classify abnormalities of the thyroid gland are Ultrasound imaging, Computer Tomography(CT),Magnetic Resonance Imaging(MRI) and Computer Aided Diagnosis (CAD).CAD help radiologists and doctors to increase the diagnosis accuracy, reduce biopsy ratio and save their time and effort. Medical image analysis has played an important role in many clinical procedures for detecting different types of human diseases. Thyroid medical images are utilized for the diagnosis process.

## IV. METHODOLOGY

### 4.1 CNN

The typical deep-learning model, Convolution Neural Networks (CNNs) can be exploited to automatically extract features from images using the hierarchical structure inspired by mammalian visual system. For image

classification tasks, traditional CNN models employ the softmax function for classification. However, owing to the limited capacity of the softmax function, there are some shortcomings of traditional CNN models in image classification. To deal with this problem, a new method combining Biomimetic Pattern Recognition (BPR) with CNNs is proposed for image classification. BPR performs class recognition by a union of geometrical cover sets in a high-dimensional feature space and therefore can overcome some disadvantages of traditional pattern recognition. The proposed method is evaluated on three famous image classification benchmarks, that is, MNIST, AR, and CIFAR-10. The classification accuracies of the proposed method for the three datasets are 99.01%, 98.40%, and 87.11%, respectively, which are much higher in comparison with the other four methods in most cases. After going through all those links let us see how to create our very own cat-vs-dog image classifier. For the dataset we will use the kaggle dataset of cat-vs-dog:

- train dataset- link
- test dataset- link

Now after getting the data set, we need to preprocess the data a bit and provide labels to each of the image given there during training the data set. To do so we can see that name of each image of training data set is either start with “cat” or “dog” so we will use that to our advantage then we use one hot encoder for machine to understand the labels(cat[1, 0] or dog[0, 1]).

```
deflabel_img(img):
word_label = img.split('.')[3]
# DIY One hot encoder
ifword_label == 'cat': return [1, 0]
elifword_label == 'dog': return [0, 1]
```

#### 4.2 SVM classifier

The standard SVM is a non-probabilistic binary classifier. binary linear classifier, i.e. it predicts, for each given input, which of two possible classes the input is a member of. or greater accuracy and kernel-function choices on low- through medium-dimensional data sets, train a binary SVM model or a multiclass error-correcting output codes (ECOC) model containing SVM binary learners using the Classification Learner app. For greater flexibility, use the command-line interface to train a binary SVM model using fitcsvm or train a multiclass ECOC model composed of binary SVM learners using fitcecoc. The overview of a technique which it's think is a very simple approach to be implemented in making comparisons with the results hyperplane formed of Support Vector Machine (SVM) on linear data to separate the two classes (binary classification), based Linear Regression method on nearest points (Closest Pair) is formed of two points between classes to take its midpoint. Obviously this method is very vulnerable to errors in the formation of the hyperplane line, because they do not have a concept of Support Vector, in the sense that all points are considered all result of midpoint or it's just take few of midpoint.

This is the linear regression formula ( $y = a + b*x$ ):

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \quad a = \frac{(\sum y - b \sum x)}{n} \quad (1)$$

#### 4.3 DECISION TREE

Decision Tree algorithm belongs to the family of supervised learning algorithms. Other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too. The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data(training data). The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.



**V. EXPERIMENTAL RESULTS**

Comparative analysis based on Silhouette Index is shown in Table 1  
Table 1

Algorithm	K=3	K=6	K=8	K=10
CNN	0.87	0.87	0.85	0.84
SVM	0.90	0.89	0.89	0.87

According to the Table 1, the proposed algorithm's performance is better than the benchmark Mean clustering. The comparative analysis of K-Means and MPI+PFFOK-Means clustering algorithms are shown in Fig 1.

**Table 2 Thyroid with Cancer Results**

Type of thyroid with cancer	Particulars	Algorithm	Type	Remarks
Female	Threshold value	Processing (CNN,SVM)	Type 1	Affected
Male	Thresholds value	Processing (CNN,SVM)	Type 2	No affected

The image chooses first and Run the main window

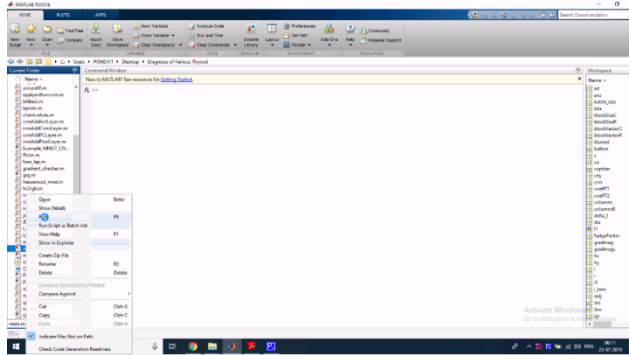


Fig: 1 Main From



Fig: 2 first image

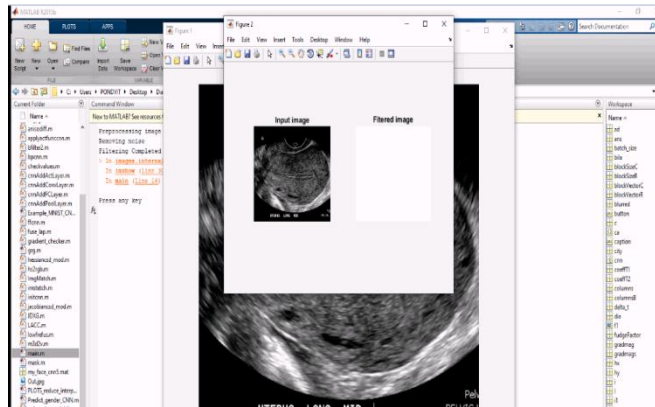


Fig: 3 FILTERED IMAGES

### VI.CONCLUSION

The application of the decision support system makes it possible to improve quality and accelerate a number of works performed. For example, the system with the use of the "HISTOLOGICAL ANALYSIS OF THYROID TUMORS" knowledge base considered in this paper makes it possible to visually compare the symptoms and diagnoses of several patients at one time. In addition, it can be used to train young physician directly during the process of diagnosis. In addition, a knowledge base can be created in according with the histological classification that is developed together with doctors that are engaged in FSBI "National Medical Research Center of Oncology named after NN Blokhin "Ministry of Health of Russia (Research Institute of Oncology named after NN Blokhin). The creation of decision support system provides the ability to store information in digital form, which is more convenient than a paper description of analyzed images. It lessens risks of information loss. The work has been done using classification data mining techniques for the diagnosis of thyroid disease. For this purpose, K nearest neighbor, Support vector machine, Decision tree and Naive Bayes classifiers have been used. The Decision Tree classifier outperformed over other classifiers. However, if we merge it with any other classification technique such as neural network, then the result might be even better as compared to what we got with the current study.

### REFERENCES

- [1].K.Saravana Kumar, Dr. R. ManickaChezian, "Support Vector Machine and K- Nearest Neighbor Based Analysis for the Prediction of Hypothyroid. International Journal of Pharma and Bio Sciences", volume - 2,Issue - 5,page no-(447-453),2014 .
- [2].<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169866/>(accessed dec 2015)
- [3].G. Zhang, L.V. Berardi, An investigation of neural networks in thyroid function diagnosis, Health Care Manage. Sci. (1998)
- [4].Xia C, Hsu W (2006) BORDER: efficient computation of boundary points. In: IEEE, 2006
- [5].Available from: <http://en.wikipedia.org>.Last accessed on Dec24.V. N. Mikhailov, G.G. Avtandilov, S.M. Zaitsev, V.G. Nikitaev, E.Yu.Berdnikovich, A.Vinogradov, I.A. Vorobiev, A.I. Pavlovskaya, N.N.Petrovichev, A. N. Pronichev, D.L. Rotin, V.A. Stepanov, ComputerSystems of histological diagnostics. Moscow Publishing House MEPI2002.
- [6] V. G. Nikitaev Modern Measurement Principles in Intellectual Systemsfor a Histological Diagnosis of Oncological Illnesses *MeasurementTechniques* 20155844674702-s2.0-8495705155610.1007/s11018-015-0736-9
- [7] V. G. Nikitaev Medical and biological measurements: ExperimentalHigh-technology information-measuring complexes of cancer diagnosis:Problems and key points of the construction methodology *MeasurementTechniques* 2015582, article no. A0172142182-s2.0-8494336213310.1007/s11018-015-0688-0
- [8] V. G. Nikitaev Methods and means of diagnostics of ontologicalDiseases on the basis of pattern recognition: Intelligent morphologicalSystems - Problems and solutions *Journal of Physics: ConferenceSeries* 201779812-s2.0-8501632035710.1088/1742-6596/798/1/012131012131.
- [9]Apte& S.M. Weiss, Data Mining with Decision Trees and Decision Rules, T.J. Watson Research Center, [http://www.research.ibm.com/dar/papers/pdf/fgcsaptewe issue\\_with\\_cover.pdf](http://www.research.ibm.com/dar/papers/pdf/fgcsaptewe issue_with_cover.pdf), (1997).
- [10].Roychowdhury S (2014) DREAM: diabetic retinopathy analysis using machine learning. In: IEEE, 2014



INNO  SPACE  
SJIF Scientific Journal Impact Factor

Impact Factor:  
7.488

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details