



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com



Prediction of Hypothyroid Disease in Early Stage Using Machine Learning Algorithms

AMULYA M B, JYOTHI T, MIZBA BANU, RANJANA G A

B.E, Student of ISE, BIET, Davangere, India

B.E, Student of ISE, BIET, Davangere, India

B.E, Student of ISE, BIET, Davangere, India

Associate Professor, Department of ISE, BIET, Davangere, India

ABSTRACT: Thyroid disease has now become the second largest disease in the endocrine system. Deep Learning for medical imaging has been on the forefront of its numerous applications, thanks to its versatility and robustness in deployment. In this project we develop the classification methodology that are employed for datasets of relatively small in size to actually train in deep learning algorithm from scratch. The aim of this work is to develop an automated thyroid diagnosis system that could aid the radiologist and fasten the diagnosis. Thyroid ultrasound datasets are classified using a small CNN from scratch, transfer learning and fine-tuning of VGG-16. We present a comparison of the aforementioned methods through accuracy and performance metrics.

KEYWORDS: CNN, VGG-16, Deep Learning, Thyroid ultrasound datasets.

I. INTRODUCTION

The thyroid gland is one of the important organs of the human body. It produces the thyroid hormone which is vital to control the body's metabolism. There are two main thyroid hormones: Triiodothyronine (T3) and Thyroxine (T4). These two hormones are mainly responsible for maintaining the energy in our bodies. Thyroid Stimulating Hormone (TSH) is produced by the pituitary gland that helps the thyroid gland to release T3 and T4. There are two common thyroid diseases- 1) Hypothyroid 2) Hyperthyroid.

Hypothyroid: When the thyroid gland cannot generate enough thyroid hormones the level of T3 and T4 becomes low and the level of TSH become high. Symptoms it presents are weight loss, tiredness, brain fog, etc.

Hyperthyroid: When the thyroid gland produces more thyroid hormone than our body actually needs, the level of T3 and T4 becomes too high and the level of TSH becomes low. Symptoms it presents are- hair loss, anxiety, sweating, etc.

Computer-aided diagnosis (CAD) systems are increasingly being used in clinical diagnostics. On the one hand, these CAD systems can reduce the drudgery of doctors. On the other hand, they can avoid some mistakes that may be made in the diagnostic process. More and more CAD systems are being applied in practice to improve the accuracy of diagnosing various diseases. Nowadays, machine learning has become an immensely popular medium for detecting various diseases. It is very convenient and effective to presume diseases using machine learning techniques. Here we have proposed a robust approach towards solving the above stated problem through the use of Machine Learning. The model proposed by us comprises of the data acquisition, Pre-processing the data followed by segmentation of the images using the convolutional neural networks to train and test and model which would help in classification as well as the localization of the disease part in the datasets of the thyroid gland. The project was tested on a dataset of the UCI machine learning dataset was used to further test the model and for future work related to thyroid ultrasound.

II. PROBLEM STATEMENT

The types of disorders that can occur because of the low or high secretion of the hormones by the thyroid gland which leads to various abnormalities and difference in shape and size. The existing system which is put in place to detect and rectify diseases related to the thyroid gland are either absolute or use the image based analysis of the gland. The



analysis of the thyroid gland and tissue is faulty in the way that it produces low resolution images and thus the subjective knowledge of the doctor is limited in analyzing the thyroid ultrasound.

III. PROPOSED SYSTEM

We have proposed a robust approach towards solving the above stated problem through the use of Machine Learning CNN method. The model proposed by us comprises of the data acquisition, Pre-processing the datasets followed by segmentation of the data and use the convolutional neural networks to train and test and model which would help in classification as well as the localisation of the disease part in the datasets of the thyroid gland.

IV. OBJECTIVES

- To design and develop a convolutional neural network that can identify thyroid disease and classify the current disease stage.
- To develop an CNN model by training the datasets collected and pre-processed.
- To provide an efficient solution for healthcare practitioners via CNN for diagnosing/classifying a particular thyroid disease that a person may have.
- To calculate the accuracy and performance of our proposed model.

V. LITERATURE SURVEY

In 2016, **K. He, X. Zhang, S. Ren, and J. Sun** present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. They explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. They provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. They also present analysis on CIFAR-10 with 100 and 1000 layers. The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, they obtain a 28% relative improvement on the COCO object detection dataset.

In 2015, **H. A. Nugroho, A. Nugroho, J. Grafika, and N. Bulaksumur** provides various segmentation methods applied on the same image like Watershed, Active contour, SVM (Support vector machine), Morphological operation etc., and give a comparison of their output. This helps the doctor to focus only on their area of interest. Key Words: Segmentation, Watershed algorithm, Active contour, Morphological Operations.

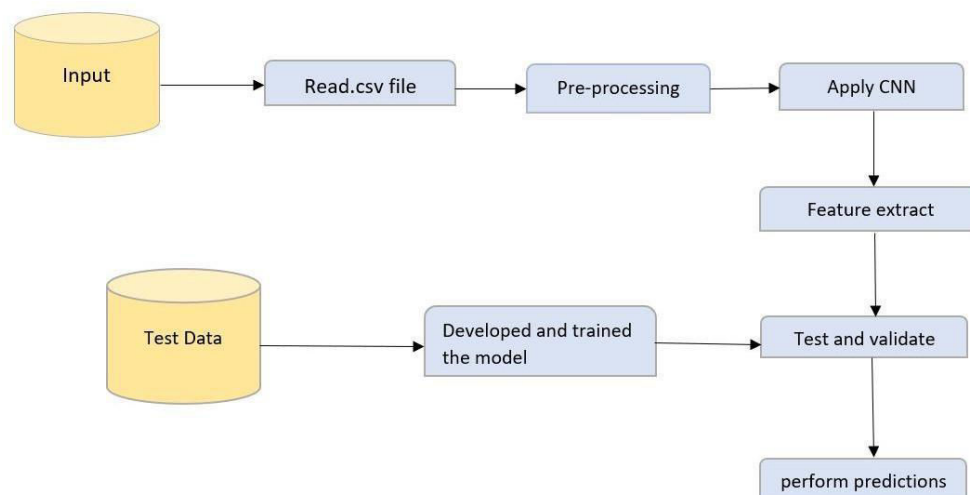
In 2016, **Hanung Adi Nugroho, Made Rahmawaty, Yuli Triyani, Igi Ardiyanto** research proposes a classification of thyroid ultrasound images by using some texture features into two classes. The dataset consists of 39 ultrasound images which grouped into 25 cystic cases and 14 solid cases. An initial step of image pre-processing is conducted to enhance the detection capability. Afterwards, followed by some methods of morphological operation, that is active contours without Edges (ACWE) and histogram equalization. The feature extraction is developed based on texture analysis by using Gray Level Cooccurrence Matrix (GLCM), Histogram and Gray Level Run Length Matrix (GLRLM). Finally, Multilayer Perceptron (MLP) is used to classify cystic nodule from solid nodule. The result shows that the proposed method achieves the accuracy of 89.74%, sensitivity of 88.89%, specificity of 91.67%, positive predictive value (PPV) of 96.00% and negative predictive value (NPV) of 78.57%. This indicates that the proposed method is excellent in classifying thyroid ultrasound images.

In 2018, **Ankita Tyagi, Ritika Mehra and Aditya Saxena** provides the analysis and classification models that are being used in the thyroid disease based on the information gathered from the dataset taken from UCI machine learning repository. It is important to ensure a decent knowledge base that can be entrenched and used as a hybrid model in solving complex learning task, such as in medical diagnosis and prognostic tasks. They also proposed different machine learning techniques and diagnosis for the prevention of thyroid. Machine Learning Algorithms, support vector machine (SVM), K-NN, Decision Trees were used to predict the estimated risk on a patient's chance of obtaining thyroid disease.

In 2019, **Naghmeh Mahmoodial, Prabal Poudel, Alfredo Illanes and Michael Friebe** a novel approach for thyroid texture characterization based on extracting features utilizing higher order spectral analysis (HOSA) was used. A Support Vector Machine (SVM) was applied on the extracted features to classify the thyroid texture. Since HOSA is a well suited technique for processing non-Gaussian data involving non-linear dynamics, good classification of thyroid

texture can be obtained in US images as they also contain non- Gaussian Speckle noise and nonlinear characteristics. A final accuracy of 93.27%, sensitivity of 0.92 and specificity of 0.62 were obtained using the proposed approach. In 2020, **Q. Zhang, J. X. Hu¹ and S. Zhou** proposed method for the detection of hyperthyroidism based on the modified LeNet-5. Three test groups were designed by randomized controlled experiment design for the detection of hyperthyroidism, in which the control group was tested manually by experienced doctors, the LeNet-5 network learning group adopted the classical LeNet-5 network learning algorithm and the experimental group used the modified LeNet-5 network learning algorithm. Evaluation indices included the accuracy, detection efficiency, sensitivity, specificity and F1 score of the detection. At the same time, differences between the two algorithms in the detection of thyroid nodules were compared. There was no significant difference between the 3 groups in the accuracy and specificity of detection of hyperthyroidism. In terms of the detection efficiency and sensitivity, the performance of the network learning algorithm group and learning group was better than that of the control group. Both the network learning algorithm group and experimental group could detect the thyroid nodules accurately, but there was no significant difference in the accuracy of detecting the type of thyroid nodules.

VI. SYSTEM DESIGN



In this project we proposed a two-track deep neural network model architecture. The first track consists of deep convolution neural network with max-pooling to enhance system ability, whereas the second track comprised of fully connected layers. The network has four layers of hidden neurons (three convolutional-pooling and one fully connected), apart from a final dense layer of output neurons (the input is not considered as a layer). The first convolution-pooling layer uses a local receptive field (convolutional kernel) of size 3×3 with a stride length of 1 pixel to extract 32 feature maps, followed by a max. pooling operation conducted in a 2×2 region, the second and third convolution-pooling layers use the same local receptive field (kernel) resulting 64 and 128 features maps respectively whereas other parameters remain unchanged. The fourth layer is fully-connected layer with 64 ReLU (rectifier linear unit) neurons and the output layers have SoftMax neurons.

Dataset Collection: Thyroid disease databases for many diseases are available freely. However, some are fully or partially open source and others are commercially available. The input dataset can be of type .csv file or clinical based on the dataset used.

Data Pre-Processing

The data pre-processing process is critical and is an essential step in data mining; it has a good effect on the data. The data pre-processing process is used to discover the data through data analysis and find the missing data; it scrutinizes the data. The preprocessing process involves data cleaning, preparation, etc., because real-world data tends to be inconsistent, noisy, and may contain missing, redundant, and irrelevant data. This may adversely affect the functioning of the algorithms. Pre-processing is used to clean the data, scale the data, and convert the data into a format compatible with the algorithms used. The pre-treatment stage in this work goes through the following stage:

- Check for missing values and if found, process them
- Check for duplicate data if you find and remove it.
- Convert all categorical features in the data set to a standardized numerical representation because of the presence of categorical and numerical features.

In the data processing stage, we identified missing data, and among these properties were missing T4 by number 151 and T3 by number 112. where we were able to Process this lost data by replacing it with the value of the mediator, and after working in this way, we were able to obtain the data in a good and better way and free from lost data, as the data became arranged and excellent and free from any defect or problem so that we can work on it smoothly and well. Data Normalization may refer to more extensive alterations aiming at harmonizing the whole probability distributions of transformed data in more complicated settings. Before averaging, normalization of ratings requires translating data from many scales to a single, apparently typical scale.

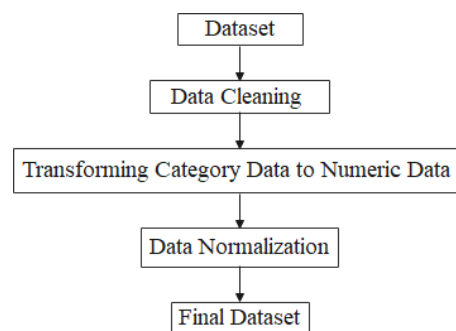


Fig: Data Pre-Processing

Feature Extraction: Running the convolutional base over your dataset, recording its output to a Numpy array on disk, and then using this data as input to a standalone, densely connected classifier similar to those you saw in part 1 of this book. This solution is fast and cheap to run, because it only requires running the convolutional base once for every input image, and the convolutional base is by far the most expensive part of the pipeline. But for the same reason, this technique won't allow you to use data augmentation.

CNN Model Evaluation: A Convolutional Neural Network (DenseNet/CNN) is a Deep Learning system that can take an input image, assign relevance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them. When compared to other classification methods, the amount of pre-processing required by a DenseNet is significantly less. While basic approaches require hand-engineering of filters, DenseNets can learn these filters/characteristics with enough training. The architecture of a DenseNet is inspired by the organization of the Visual Cortex and is akin to the connectivity pattern of Neurons in the Human Brain. Individual neurons can only respond to stimuli in a small area of the visual field called the Receptive Field. A number of similar fields can be stacked on top of each other to span the full visual field.

Working of CNN

A Convolutional Neural Network has three layers in general.

- Input: If the image has 32 widths and 32 heights and three R, G, and B channels, it will retain the image's raw pixel([32x32x3]) values.
- Convolution: It computes the output of those neurons that are related to input's local regions, such that each neuron calculates a dot product between weights and a small region in the input volume to which they are actually linked. If we opt to use 12 filters, for example, the volume will be [32x32x12].
- Layer of ReLU: It's utilized to apply an elementwise activation function, such as $\max(0, x)$ thresholding at zero. It yields ([32x32x12]), which corresponds to a volume of the same size.
- Pooling: This layer is used to perform a down-sampling operation along the spatial dimensions, resulting in a volume of [16x16x12].

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

Convolved Feature

Fig : Feature Map

- Locally Connected: It is a normal neural network layer that accepts an input from the preceding layer, computes the class scores, and outputs a 1-Dimensional array with the same size as the number of classes.
- Next, we'll apply a Pooling layer to our Convolutional layer, so that we may construct a Pooled feature map from each feature map. The pooling layer's major goal is to ensure that our images exhibit spatial invariance. It also aids in reducing the size of our photos and preventing overfitting of our data. We'll then flatten all of our pooled photos into a single long vector or column containing all of these values, which we'll then feed into our artificial neural network. Finally, to get the final output, we'll feed it into the locally connected layer.

Fig : Fully Connected Layer

Classification: Classification is a supervised learning approach for machine learning task. It requires labeled dataset to map the data into specific groups or classes. There are various classification algorithms used to classify the thyroid disease data set such as support vector machine, feed forward neural network, back propagation neural network, k- nearest neighbor, decision trees, etc. We have used CNN classification.

```

: model = Sequential()
model.add(Dense(256, input_shape=[x.shape[1]], activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.2))

model.add(Dense(1, activation='sigmoid'))

: model.summary()

Model: "sequential"
Layer (type)                Output Shape              Param #
-----
dense (Dense)                (None, 256)               7424
dropout (Dropout)            (None, 256)               0
dense_1 (Dense)              (None, 128)               32896
dropout_1 (Dropout)          (None, 128)               0
dense_2 (Dense)              (None, 63)                8127
dropout_2 (Dropout)          (None, 63)                0
dense_3 (Dense)              (None, 1)                 64
Total params: 48,511
Trainable params: 48,511
Non-trainable params: 0

```

Fig : CNN Model Summary

VII. RESULT

In the PyCharm, user can view the predictFromModel.py code, and also have flask app that provides useful tools and features for model, user will receive the path through app.py which is the generic entry point for abseil python applications. The link is received is of http format: <http://127.0.0.1:5000>. A user can received this page by using

<http://127.0.0.1:5000> link path, where the .csv file can be uploaded which was already saved. On uploading the .csv file user can receive the required result for prediction which can later be downloaded by using download option. A user can received result for prediction is of the above format. The prediction is mainly based on three categories: Primary hypothyroid, Compensated hypothyroid, negative. The primary hypothyroid stands for low levels of blood thyroid hormone due to destruction of thyroid gland. Compensated hypothyroid stands for the combination of elevated serum thyrotropin and normal serum thyroxine and negative stands for the no thyroid antibodies were found.

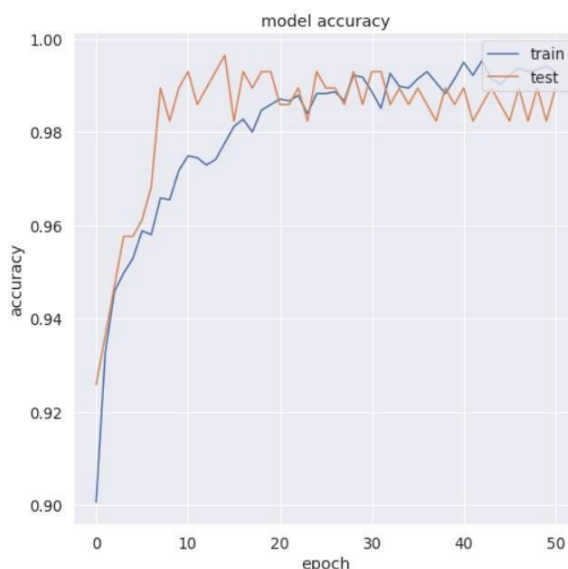


Fig: Model Accuracy

Our model gives the training and testing accuracy above 98%.

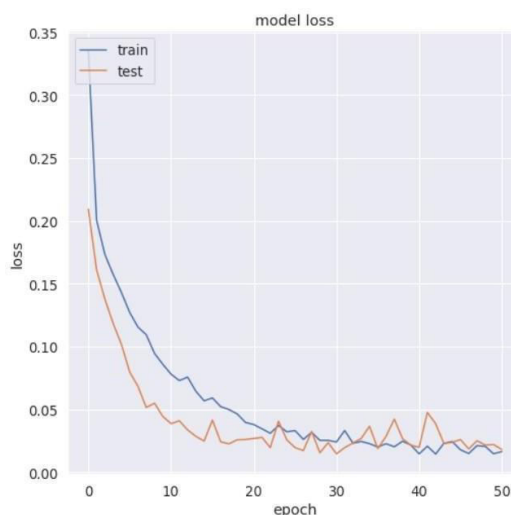


Fig : Model Loss

The above graph illustrates the training loss and accuracy loss of the model.



VIII. CONCLUSION

A convolutional neural network with optimization-based computer-aided diagnosis of thyroid diseases will be developed. A modified Dense Net architecture of convolutional neural network is employed, and the training method is improved. The architecture is modified by adding the trainable weight parameters to each skip connection in Dense Net. And the training method is improved by optimizing the learning rate with flower pollination algorithm for network training. Experimental results can demonstrate that the proposed method of convolutional neural network is efficient for the diagnosis of thyroid diseases, and it has superior performance than other CNN methods.

REFERENCES

1. "Thyroid Nodule Segmentation Using Active Contour Bilateral Filtering on Ultrasound Images", H. A. Nugroho, A. Nugroho, J. Grafika, and N. Bulaksumur, pp. 43–46, 2015.
2. "Deep residual learning for image recognition," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), K. He, X. Zhang, S. Ren, and J. Sun pp. 770–778, Las Vegas, NV, USA, June-July 2016.
3. "Texture Analysis for Classification of Thyroid Ultrasound Images," Hanung Adi Nugroho, Made Rahmawaty, Yuli Triyani, Igi Ardiyanto, IEEE-2016.
4. "Interactive Thyroid Disease Prediction System Using Machine Learning Technique," Ankita Tyagi, Ritika Mehra and Aditya Saxena IEEE-2018.
5. "Higher Order Statistical Analysis for Thyroid Texture Classification and Segmentation in 2D ultrasound Images," Naghmeh Mahmoodial, Prabal Poudel, Alfredo Illanes and Michael Friebe IEEE-2019.
6. "The Detection of Hyperthyroidism by the Modified LeNet-5 Network," published in "Biomedical Research in Healthcare Setting" Q. Zhang, J. X. Hu1 and S. Zhou, Indian J Pharm Sci 2020:82(3)Spl issue5;108-114.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details