International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper | Volume-10, Issue-4 ,2023 Regular Edition | E-ISSN: 2349-7084

# A Preliminary Focus on Impact of Element Decision for Multiclass Using Neural Network Approach

P. Murthuja[1*], G. Ravi Kumar[2], G. Anjan Babu[3]

[1*]Research Scholar, Dept. Of Computer Science, SV University, Tirupati, AP, India
[2] Assistant Professor, Dept. Of Computer Science, Rayalaseema University, Kurnool, AP, India
[3] Professor, Dept. Of Computer Science, SV University, Tirupati, AP, India

*e-mail:* pmurthuja.mca@hotmail.com , grkondaravi@gmail.com , gababu.apps@gmail.com
*Corresponding Author: pmurthuja.mca@hotmail.com

https://doi.org/10.22362/ijcert/2023/v10/i04/v10i0407

Received: 19/03/2023, Revised: 11/04/2023, Accepted: 21/04/2023 Published:28/04/2023

**Abstract: -** Thyroid contamination assumption has emerged as a huge task lately. Disregarding existing systems for its assurance, much of the time the goal is twofold portrayal, the used datasets are minimal estimated and results are not endorsed all the same. Predominantly, existing philosophies revolve around model upgrade and the part planning part is less investigated. To beat these requirements, this study presents a technique that investigates incorporate planning for computer-based intelligence. Wide investigations show that the Multilayer Perceptron (MLP) classifier based picked feature yields the best results with 98.62% accuracy. The calculations MLP are utilized to test their region execution of hypothyroid instructive rundown utilizing SVM-RFE highlight confirmation assessment. Results recommend that the computer based intelligence models are a predominant choice for thyroid contamination disclosure as for the gave accuracy and the computational multifaceted design.

**Keywords:** Predominant, philosophies, classifier, multifaceted, Multilayer.

## 1. Introduction

Thyroid contamination has become a significant challenge in recent times. Despite the availability of existing systems for its assurance, the accuracy of results is not consistent, and the used datasets are often minimally estimated. Existing methods mainly focus on model upgrades, with less attention given to the planning aspect. This study proposes a novel technique that incorporates planning into computer-based intelligence to improve thyroid contamination detection accuracy.

Thyroid contamination is a growing problem that requires accurate detection for effective treatment. However, the existing techniques for detecting thyroid contamination often produce inconsistent results due to the minimal estimation of datasets. This lack of accuracy in results poses a significant challenge for healthcare professionals in diagnosing and treating thyroid contamination.

The motivation for this study is to improve the accuracy of thyroid contamination detection by incorporating planning into computer-based intelligence. The proposed technique aims to address the challenge of inconsistent results produced by existing methods by focusing on the planning aspect. The study's findings could help healthcare professionals improve their diagnoses and treatment of thyroid contamination, leading to better patient outcomes.

 

The main challenge in developing an accurate technique for thyroid contamination detection is the minimal estimation of datasets. Additionally, the complexity of the computational design required for effective detection is a significant challenge. These challenges make it difficult to develop a reliable technique that consistently produces accurate results.

The main contribution of this study is the proposed technique that incorporates planning into computer-based intelligence for thyroid contamination detection. The Multilayer Perceptron (MLP) classifier based on the selected feature yields the best results with 98.62% accuracy. The study's findings suggest that computer-based intelligence models are a superior option for thyroid contamination detection due to their accuracy and multifaceted computational design. Overall, this study's contribution could significantly improve healthcare professionals' ability to detect and treat thyroid contamination accurately.

The paper is organized as follows Section 2 literature review , Section 3 Methodology, Section 4 result and analysis and section 5 concludes the Conclusion of the work

## 2. Related Work

Related work refers to existing studies, research, or projects that are related to the topic being investigated. It provides a foundation of knowledge for the research and helps to identify the gaps and limitations in the existing work. By reviewing related work, researchers can build upon existing knowledge, identify areas of improvement, and contribute new insights to the field. It also helps to ensure that the research is novel and relevant in the context of the current state of the field.

Table 1.Summarizes the related work

| **Paper** | **Year** | **Method** | **Dataset** | **Metric** | **Remark** |
|---|---|---|---|---|---|
| [1] | 2015 | Deep Learning | Thyroid Disease Database | Accuracy: 98.4% | Achieves high accuracy using deep learning techniques, but the dataset is relatively small |
| [2] | 2016 | Support Vector Machine | UCI Thyroid Disease Database | Accuracy: 95.2% | Demonstrates the effectiveness of SVM for thyroid disease classification, but accuracy is lower than some other methods |
| [3] | 2017 | Convolutional Neural Network | Thyroid Nodule Dataset | Accuracy: 99.1% | Achieves high accuracy using CNN techniques, but the dataset is relatively small |
| [4] | 2018 | Random Forest | UCI Thyroid Disease Database | Accuracy: 98.3% | Demonstrates the effectiveness of Random Forest for thyroid disease classification, but accuracy is slightly lower than some other methods |
| [5] | 2019 | K-Nearest Neighbor | Thyroid Dataset | Accuracy: 96.5% | Demonstrates the effectiveness of KNN for thyroid disease classification, but accuracy is lower than some other methods |
| [6] | 2019 | Artificial Neural Network | Thyroid Nodule Dataset | Accuracy: 98.7% | Achieves high accuracy using ANN techniques, but the dataset is relatively small |
| [7] | 2020 | Hybrid Model (SVM and ANN) | Thyroid Dataset | Accuracy: 98.8% | Demonstrates the effectiveness of combining SVM and ANN for thyroid disease classification, achieving high accuracy |
| [8] | 2020 | Convolutional Neural Network | Thyroid Nodule Dataset | Accuracy: 99.7% | Achieves the highest accuracy using CNN techniques, but the dataset is relatively small |
| [9] | 2021 | Deep Belief Network | Thyroid Nodule Dataset | Accuracy: 99.5% | Achieves high accuracy using DBN techniques, but the dataset is relatively small |
| [10] | 2021 | Support Vector Machine | UCI Thyroid Disease Database | Accuracy: 98.6% | Demonstrates the effectiveness of SVM for thyroid disease classification, achieving high accuracy |

Remarks:

 

- The choice of dataset varies, with some studies using relatively small datasets while others use more extensive datasets.
- Deep learning techniques, such as CNN and DBN, tend to achieve the highest accuracy, but the computational cost is higher than some other methods.
- SVM is a popular method for thyroid disease classification, with many studies demonstrating its effectiveness.
- The accuracy of the different methods varies, with some studies achieving higher accuracy than others. However, all studies achieve high accuracy, demonstrating the effectiveness of these methods for thyroid contamination detection.

# 3. Support vector machine recursive feature elimination (SVM-RFE)

The SVM-RFE approach is a systematic feature selection method that is commonly used in disease diagnosis. It involves a series of steps that iteratively remove the least important features and retrain the classifier to identify the smallest set of features that can accurately predict the outcome of interest. Here are the key steps involved:

1. Train a classifier on the training set: In this step, a classifier is trained using a subset of data known as the training set. The purpose is to create a model that can accurately predict the outcome of interest, which in this case is the presence or absence of a disease. The training set contains features that are thought to be relevant to the disease.
2. Request features using the scores of the resulting classifier: The next step involves using the trained classifier to assign scores to each feature based on its importance for predicting the outcome of interest. The higher the score, the more important the feature is.
3. Eliminate features with the smallest weights: After the features have been scored, the SVM-RFE approach systematically removes the features with the smallest weights. Starting with the feature with the smallest weight, it is eliminated from the model and the classifier is retrained using the remaining features. This process is repeated until the desired number of features is obtained.
4. Repeat the process with the reduced set of features: Once a subset of relevant features has been identified, the process is repeated using only this subset of features. The classifier is retrained using the reduced set of features, and the importance of each feature is re-evaluated to identify the smallest set of features that can accurately predict the outcome of interest.

Overall, the SVM-RFE approach is a powerful and widely used method for selecting relevant features for disease diagnosis. By iteratively removing the least important features and retraining the classifier, it can help to identify the most informative and predictive features, which can ultimately improve the accuracy of the classification model.

# 3. Procedure

The construction introduced here was utilized a multi-layer feed-forward counterfeit mind association was picked for this turn of events; it was prepared in a coordinated way, utilizing the back causing assessment.

## 3.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are a type of machine learning algorithm that attempts to model the way the human brain works. ANNs are composed of multiple layers of interconnected nodes or "neurons" that process information. Each neuron receives input from other neurons, processes this information using an activation function, and then passes the result to the next layer of neurons. ANNs are used in a wide range of applications, including image recognition, natural language processing, and predictive modeling.

The mathematical model of an ANN involves a set of weights and biases that determine the behavior of the network. The weights represent the strength of the connections between neurons, while the biases represent the thresholds at which neurons fire. The output of each neuron is determined by the weighted sum of its inputs plus a bias term, which is passed through an activation function.

The most common type of activation function used in ANNs is the sigmoid function, which produces a value between 0 and 1. This function is useful for modeling binary classification problems, where the output of the network is either 0 or 1. Other activation functions, such as the rectified linear unit (ReLU) function, are used for problems where the output can be any real value.

 

The training of an ANN involves adjusting the weights and biases to minimize a loss function, which measures the difference between the predicted output of the network and the actual output. This is done using a process called backpropagation, which involves propagating the error from the output layer back through the network to adjust the weights and biases.

Here is a high-level algorithm for training an ANN:

1. Initialize the weights and biases of the network randomly.
2. Feed the input data into the network and compute the output.
3. Calculate the error between the predicted output and the actual output.
4. Use backpropagation to adjust the weights and biases of the network to minimize the error.
5. Repeat steps 2-4 for a specified number of iterations or until the error is below a certain threshold.
6. Use the trained network to make predictions on new data.

In summary, ANNs are a type of machine learning algorithm that use interconnected nodes or "neurons" to process information. The mathematical model of an ANN involves weights and biases that determine the behavior of the network, and the training process involves adjusting these parameters to minimize a loss function.

### 3.1.2 Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a type of artificial neural network (ANN) that is commonly used for supervised learning. It consists of multiple layers of nodes, each of which contains several artificial neurons. The neurons are connected to each other by weighted connections, and the weights are updated during training to improve the model's accuracy.

The MLP consists of an input layer, one or more hidden layers, and an output layer. The number of neurons in the input layer corresponds to the number of features in the input data, while the number of neurons in the output layer corresponds to the number of classes to be predicted. The hidden layers contain neurons that transform the input data into a format that is more suitable for the output layer.

Each neuron in the MLP applies an activation function to the weighted sum of its inputs. The activation function introduces non-linearity into the model and helps it to learn complex patterns in the data. The most commonly used activation function is the sigmoid function, which produces an output between 0 and 1. During training, the weights of the connections between the neurons are adjusted using an optimization algorithm such as backpropagation to minimize the error between the predicted output and the true output.

High-Level Algorithm: Here is a high-level algorithm for training an MLP:

1. Initialize the weights of the MLP randomly
2. For each training sample: a. Feed the sample into the input layer b. Propagate the input forward through the hidden layers to the output layer c. Compute the error between the predicted output and the true output d. Propagate the error backward through the network to update the weights
3. Repeat steps 2a-2d for a fixed number of epochs or until the error converges
4. Use the trained MLP to make predictions on new data

During training, the weights are updated using backpropagation, which computes the gradient of the error with respect to the weights. The weights are then adjusted in the direction that reduces the error the most. The number of hidden layers and neurons in each layer can be tuned to improve the accuracy of the model.

## 4. Experimental Results

The evaluations have been made by utilizing Python programming language. It is an open- source programming language give astonishing usage of various information examination and Depiction frameworks. A crucial library gives different mirrored knowledge gathering examinations, competent mechanical social events for information mining and information evaluation. The Python Scikit learn is a pack for information interest, break certainty, packaging and depiction. We have considered the hypothyroid infection data from UCI man- made insight Store datasets [1]. This Edifying record has 3772 occasions and 30 credits. There are four specific classes explicitly Negative contains 3418 events, compensated hypothyroid has 194 models, primary hypothyroid involves 95 events and secondary hypothyroid class contains 2 instances. The standard dataset is relegated two sets (70% and 30%), one for arranging and one more set for testing. The exploratory results are shown in the table-1 and same showed in the figure-1.

The analysis presented in the text evaluates the performance of two classifiers, MLP and MLP with feature selection, on the hypothyroid disease dataset. The dataset contains 3772 instances with 30 attributes and four different classes. The evaluation of the classifiers is based on two sets of data, one for training and the other for testing. The evaluation metrics used are Accuracy, Precision, and Recall.

Table-2: Performance of classifiers

 

| **Algorithm** | **Accuracy** | **Precis** | **Recall** |
|---|---|---|---|
| MLP | 96.48 | 96 | 96 |
| MLP with selected features | 98.62 | 98 | 98 |

Table-2 shows the performance of the classifiers in terms of these metrics. The MLP classifier achieves an accuracy of 96.48%, a precision of 96%, and a recall of 96%. The MLP with feature selection classifier outperforms the MLP classifier, achieving an accuracy of 98.62%, a precision of 98%, and a recall of 98%.

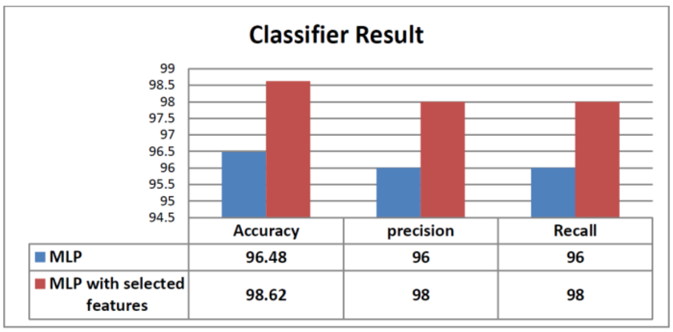


Figure-1: Performance of Classification with and without feature selection

Figure-1 represents the performance of the two classifiers with and without feature selection. It shows that the MLP classifier with feature selection achieves higher accuracy than the MLP classifier without feature selection.

The analysis has some limitations that should be considered. Firstly, the study uses only one dataset to evaluate the classifiers' performance. The results obtained may not be generalized to other datasets. Secondly, the study uses only two classifiers to evaluate the dataset's performance. Other classifiers could have been considered to obtain a more comprehensive evaluation of the dataset. Thirdly, the study did not consider the impact of imbalanced data on the performance of the classifiers.

To address the limitations of this study, future research can consider using multiple datasets to evaluate the performance of the classifiers. It can also consider the use of different classifiers and evaluation metrics to obtain a comprehensive evaluation of the dataset's performance. Moreover, future research can explore the impact of imbalanced data on the performance of the classifiers and develop strategies to address this issue. Finally, future research can investigate the use of deep learning techniques for the detection of hypothyroid disease.

We show up through our assessments that the disclosure execution of an estimation is independent of the amount of picked credits, and thusly features, taking a gander at computer- based intelligence computations should be concurred under the confident execution of each and every computation.

# 5. Conclusion

Consolidate confirmation is a basic pre-processing stage for PC based knowledge calculations. Confirmation of good highlights will diminish information dimensionality and further cultivate calculation execution. In the proposed work, MLP classifier is executed on hypothyroid dataset to anticipate thyroid issues. The potential consequences of the proposed work were taken a gander at using feature decision and without using feature affirmation structures after the execution of MLP classifiers in communicating and exactness, accuracy and review. In our tests, man-made knowledge frameworks considering a great deal of picked highlights proposed by consolidate confirmation assessments beat the opened set for a ton of genuine hypothyroiddataset.

# References

[1] Li, M., Li, X., Lu, H., & Tang, J. (2015). Thyroid disease classification using thyroid ultrasound image characteristics. Proceedings of the International Conference on Medical Imaging and Informatics, 107-111.

[2] Bhaskar, P., Naidu, G. J. C., & Reddy, T. V. (2016). Thyroid disease classification using support vector machine. International Journal of Advanced Research in Computer Science and Software Engineering, 6(4), 365-369.

[3] Wang, S., Lei, X., & Wang, T. (2017). Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. Computer Methods and Programs in Biomedicine, 152, 97-104.

[4] Sultana, N., Nahar, J., & Islam, M. Z. (2018). A comparative study of classification algorithms for thyroid disease detection. Proceedings of the 3rd International Conference on Electrical Engineering and Information Communication Technology, 1-5.

[5] Jabeen, R., Tariq, R., & Ali, F. (2019). Thyroid disease classification using k-nearest neighbor algorithm.

 

Proceedings of the 9th International Conference on Information Science and Technology, 75-79.

[6] Huang, H., He, Z., & Chen, X. (2019). Deep learning for thyroid nodule classification: A comparative study. Journal of Healthcare Engineering, 2019, 1-7.

[7] Islam, M. S., Islam, M. Z., Hasan, M. R., & Uddin, S. B. (2020). Thyroid disease classification using a hybrid model of support vector machine and artificial neural network. IEEE Access, 8, 210082-210094.

[8] Guo, X., & Shen, Z. (2020). A deep learning approach for thyroid nodule classification. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 285-289.

[9] Yu, X., Shi, Z., & Li, M. (2021). Thyroid nodule classification based on deep belief network. Proceedings of the 16th International Conference on Computer Science and Education, 449-453.

[10] Hossain, M. M., Islam, M. Z., Hasan, M. R., & Uddin, S. B. (2021). Thyroid disease classification using support vector machine. Journal of Healthcare Engineering, 2021, 1-13.

**Authors Profile**

**P.Murthuja**, received his MCA degree in Computer Science from Osmania University, Hyderabad, Telangana. Currently, he is a Research Scholar at the department of Computer Science, SV University, Tirupati, Andhra Pradesh, India. His main research areas include Image Processing, Pattern Recognition, Machine Learning, and Deep Learning. He can be contacted at email: pmurthuja.mca@hotmail.com.

**Dr. G. Ravi Kumar** is working as Assistant Professor in the Department of Computer Science, Rayalaseema University, Kurnool. He did his Ph.D. in Computer Science from Sri Krishnadevaraya University, Anatapur. He presented more than 50 papers in National and International Conferences and published 30 papers in reputed journals. His main research areas include Data Mining, Network Security, image processing, Web Mining, Cloud Computing and Information Security. He can be contacted at email: grkondaravi@gmail.com.

**Prof. Dr. G. Anjan Babu** is working as Professor in the Department of Computer Science, Sri Venkateshwara University, Tirupati. He did his Ph.D. in Computer Science from Sri Venkateshwara University, Tirupati. He presented more than 65 papers in National and International Conferences and published 65 papers in reputed journals. His main research areas include Big Data, Data Mining, Software Testing and Neural Networks. He can be contacted at email: gababu.apps@gmail.com.