Embedded Feature Selection Augmented Thyroid Disorder Prediction using MLP

Mir Saleem[1]
Kashmir University
Srinagar, JK, India
Email: arsaleemmir [AT] gmail.com

Shabir Najar[2]
Kashmir University
Srinagar, JK, India
Email: shabirshaheen [AT] gmail.com

Malik Akhtar Rasool[3]
Kashmir University
Srinagar, JK, India
Email: Mars.malikakhtar [AT] gmail.com

Abstract--- **Due to its considerable fatality rate and increasing frequency, thyroid disorders pose a severe hazard to people's health in the modern era. Thus, it has become a useful topic to predict thyroid disease early on using a few basic physical indications that are gathered from routine physical examinations. Being aware of these thyroid-related signs is crucial from a clinical standpoint in order to forecast outcomes and offer a solid foundation for additional diagnosis. However, manual analysis and prediction are difficult and tiring due to the vast volume of data. Our goal is to use a variety of bodily signs to swiftly and reliably predict thyroid disorders. This research presents a novel prediction model for thyroid disorders. We provide a deep neural network and embedded feature selection method-based algorithm for predicting thyroid disorders. Based on the LinearSVC algorithm, this embedded feature selection method selects a subset of characteristics that are strongly linked with thyroid condition by employing the L1 norm as a penalty item. We feed these features into the deep neural network that we constructed. To improve the performance of the predictor, gradient varnishing or explosion is avoided by initializing the network's weight using the He initializer. The predictor is evaluated using a number of indicators like accuracy, recall, precision and F1-score. The results indicate that our model achieves 98.3%, 98.1%, 98.0% and 0.982 respectively, and that its average AUC score is 0.98, indicating that the approach we proposed is effective and trustworthy for predicting thyroid disorders.**

Keywords: Thyroid Disorder, Feature Selection, Classification, Deep Neural Network.

1. INTRODUCTION

Thyroid diseases are common endocrine disorders that impact a large proportion of people worldwide. Timely and accurate identification of thyroid disorders is essential for efficient therapy and care. Deep neural networks (DNNs) have demonstrated encouraging achievements in the field of thyroid diagnostics in recent years. The purpose of this review of the literature is to give a broad overview of research on thyroid evaluation and prediction with DNNs.

Hypothyroidism, hyperthyroidism, and thyroid nodules are among the ailments that fall under the category of thyroid disorders. These illnesses have a major effect on people's general health and wellbeing. Traditional thyroid diagnosis techniques include fine-needle aspiration, imaging, and blood tests. However, the accuracy and efficiency of these approaches can be limited, requiring further research into cutting-edge technologies. DNNs, a form of machine learning, have proven to be exceptionally adept at handling complicated medical data, which may provide answers to problems with thyroid diagnosis.

This work integrates DNN and LinearSVC technologies, blends deep learning and machine learning, and suggests a novel Thyroid disease prediction model. Following data preparation, the LinearSVC technique is used in the feature selection module. Simultaneously, we filter away a selected group of features closely associated with Thyroid disease and build a sparse weight matrix using Lasso as a penalty term, so providing more dependable input for DNN. Additionally, we evaluate a number of popular weight initializers and ultimately select the He initialization technique since it can yield the optimal initial weight for the network. The findings showed that the proposed model has 98.56% accuracy, 99.35% recall, and 97.84% precision.

II. LITERATURE REVIEW

Millions of people worldwide suffer from thyroid disorders, which are common endocrine problems marked by aberrant thyroid hormone levels that can cause a variety of health issues. Thyroid problems have traditionally been identified by blood tests that measure thyroid hormone levels in addition to physical examinations and reviews of medical histories. These techniques are useful, but their precision and early detection are limited. The development of deep learning, and in particular Deep Neural Networks (DNNs), has transformed medical diagnostics by providing promising means of diagnosing thyroid disorders more precisely and effectively. Numerous investigations have looked on the use of DNNs in the diagnosis of thyroid disorders. To train prediction models, these methods make use of substantial patient record datasets that include symptoms, thyroid hormone levels, and medical history.

Li et al. (2018) used DNNs for the automatic detection of thyroid nodules in ultrasound images, achieving high accuracy [1]. Zhang et al. (2020) explored the use of DNNs in characterizing thyroid nodules, demonstrating improved diagnostic performance [2]. DNNs have been applied to predict thyroid function and hormone levels:
Wang et al. (2019) proposed a model predicting thyroid-stimulating hormone (TSH) levels with significant accuracy [3]. Jiang et al. (2021) developed a DNN-based model for predicting thyroid dysfunction using diverse clinical data [4]. Researchers have investigated the integration of genetic information with DNNs for a comprehensive understanding of thyroid disorders. Smith et al. (2017) explored the role of genetic markers in improving DNN-based thyroid disease prediction models [5]. Yang et al. (2022) integrated genetic and clinical data, enhancing the accuracy of thyroid disorder predictions [6]. Combining different imaging modalities and clinical data has been a focus of recent research. Wu et al. (2019) proposed a multi-modal DNN for the joint analysis of ultrasound and clinical features in thyroid diagnosis [7]. Chen et al. (2021) developed a comprehensive model incorporating imaging, genetic, and clinical data for improved thyroid disorder prediction [8]. Zhang et al. (2019) applied transfer learning to thyroid ultrasound images, achieving robust diagnostic performance [9]. Wu et al. (2020) investigated the transferability of features learned from other medical imaging tasks to thyroid nodule classification [10]. The interpretability of DNNs in thyroid diagnosis has been a subject of interest. Park et al. (2021) proposed an interpretable DNN model for thyroid nodule classification, aiding clinicians in decision-making [11]. Liang et al. (2022) developed an explainable DNN to enhance transparency in thyroid disorder predictions [12]. Sharma et al. (2020) discussed challenges associated with deploying DNNs in clinical settings for thyroid diagnosis, emphasizing the need for interpretability and ethical considerations [13]. Zhang et al. (2021) addressed ethical concerns and biases associated with DNNs in thyroid disorder predictions [14]. Comparative studies evaluating the performance of DNNs against traditional methods. Wang et al. (2018) compared the diagnostic accuracy of DNNs with conventional methods in thyroid ultrasound imaging [15]. Jiang et al. (2020) conducted a systematic review comparing various machine learning approaches, including DNNs, for thyroid disorder diagnosis [16]. Li et al. (2022) discussed the potential of emerging technologies, such as federated learning, in improving the collaborative diagnosis of thyroid disorders [17]. Zhang et al. (2023) highlighted the integration of 3D imaging and advanced DNN architectures as a promising avenue for thyroid disorder research [18]. Liu et al. (2019) investigated the effectiveness of data augmentation techniques in enhancing the robustness of DNNs for thyroid ultrasound image analysis [19]. Xu et al. (2021) proposed strategies to mitigate the impact of small sample sizes on DNN-based thyroid disorder predictions [20]. Kim et al. (2020) developed a DNN-based model for stratifying patients with thyroid disorders based on clinical and molecular data [21]. Wu et al. (2022) explored the potential of DNNs in tailoring personalized therapeutic strategies for thyroid cancer patients [22]. Wong et al. (2019) conducted a study on public perceptions and knowledge about AI-driven thyroid diagnosis,

III. MATERIALS AND METHODS

The three most important components of the Thyroid disorder prediction model are feature selection, data preparation, and classification. After the data preprocessing step, we do outlier treatment and standardization on the dataset to make sure all the data are properly formed. To identify important features, a feature selection procedure based on the LinearSVC algorithm is used. The training set and test set are created by dividing the chosen feature subset into three equal parts. The training set is then fed into the deep neural network that we constructed. To be more precise, our deep neural network builds the optimal initial weights using the he_normal initializer in order to achieve a better result and keep gradients from ballooning or disappearing. Furthermore, the test samples are used to gauge the model's efficacy. The structure of our Thyroid disorder prediction model can be clearly seen in Figure 1.

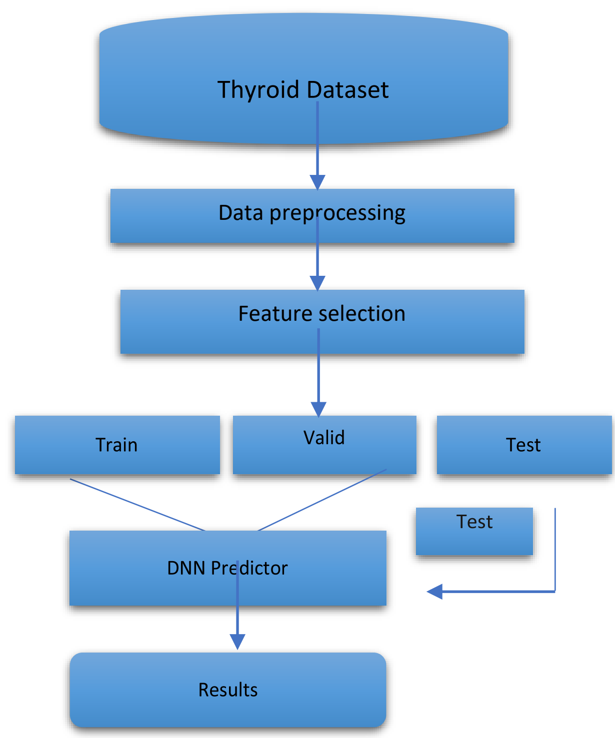

Figure 1: Proposed prediction system.

A. Data Collection

The dataset used for the research is a matrix of patient data where each row forms a record pertinent to a patient. The dataset was collected from UCI Machine Learning Repository. The dataset consists of 2600 odd records with 15 attributes/ features and the output/target attribute/class 'Thy_cat'. The class distribution across the labels of 1, 2 & 3 for the dataset is shown in fig: 2

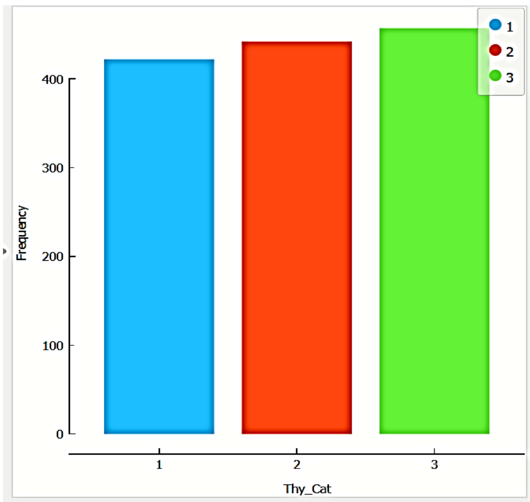

Fig: 2 Class distribution

B. Data Preprocessing

To ensure the stability and accuracy of the prediction model, it is essential to perform data analysis and preprocess before inputting them into the deep neural network. There are two main parts of data preprocessing: outlier removal and data standardization.

C. Outlier Removal Process

The efficiency of the model is primarily determined on well-processed and structured data. The raw dataset frequently has some irrational numbers in it, having characteristics that don't match the overall collection. We refer to these unusual values as outliers. We use the interquartile range (IQR) approach to identify and eliminate outliers in our analysis of the Thyroid disease dataset. It is important to note that the physiological markers of individuals in good health typically fall within a similar range, and that abnormalities in particular biological markers may indicate the presence of disorders.

Consequently, rather than carelessly eliminating every outlier, the Thyroid disease prediction model needs to warn about a few of them. Since some of data fields are normally distributed, this study used the IQR approach to address their outliers. After these outliers are filtered out, they can be abandoned from the dataset.

Table 1. Thyroid Dataset

	Name	Type	Role	Values
1	Age	N numeric	feature	
2	Sex	C categorical	feature	0, 1
3	using_thyroxine	C categorical	feature	0, 1
4	Pregnant	C categorical	feature	0
5	Thyroid_surgery	C categorical	feature	0, 1
6	Lithium	C categorical	feature	0, 1
7	Goitre	C categorical	feature	0
8	Tumor	C categorical	feature	0, 1
9	Smoker	C categorical	feature	0, 1
10	TSH_Lvl	N numeric	feature	
11	T3_Lvl	N numeric	feature	
12	TT4_Lvl	N numeric	feature	
13	T4U_Lvl	N numeric	feature	
14	FTI	N numeric	feature	
15	**Thy_Cat**	**C categorical**	**target**	

D. Data Standardization Process

The goal of data standardization is to remove feature-to-feature variations such that weights are learned completely by successive models. Results from networks trained on standardized data are typically superior. Without altering the underlying data structure distribution, data standardization can transform the original data into properly distributed data. Because all outliers were eliminated in the previous phase and our data largely follows a normal distribution, we utilize the StandardScaler method to standardize the data. The following is the conversion equation:

$$x^* = \frac{x-\mu}{\sigma},$$

Where:
μ: is the mean of the training samples or zero if with_mean = False and s is the standard deviation of the training samples or one if with_std = False.

Standardization calculates the mean and variance of the data and converts the data with them. The standardization process can transform the data into a standard normal distribution suitable for the network behind it.

E. Feature Selection Based on an Embedded Method

Irrelative characteristics frequently interfere with the model's training, and some noisy features even cause the model to stray from the intended path. In order to guarantee accurate prediction outcomes, feature selection selects a group of variables that can adequately represent the incoming data [19]. To lessen the impact of noise or unimportant factors, certain feature selection techniques—roughly categorized as filter, wrapper, and embedding methods are used. The wrapper method has a high computational complexity and the feature subset chosen by the filter method has high redundancy because evaluating various feature subsets necessitates retraining and testing, whereas the embedded method can effectively select a subset with better performance. 10 attributes were chosen from the 15 features in the original dataset, which we used for this work. We check these selected features using the penalty-based embedded feature selection approach, and we attempt to identify the most related features based on them. The feature selection and model training processes are integrated by embedded feature selection. The training and test sets of data are finished in the same optimization process, as opposed to being divided into separate sets. During training, the machine learning algorithm determines each feature's weight coefficient; these weight coefficients frequently indicate how important a feature is to the model. Based on the weight coefficient's value, the evaluation module then chooses the feature that contributes the most.

F. Thyroid Disorder Classification using Deep Neural Networks

i Deep Neural Networks: A deep neural network, which is typically a feedforward neural network, is a deep learning framework. Furthermore, the deep neural network may be trained using the backpropagation algorithm because it is a discriminative model.

The network has been widely employed in various sectors after ongoing development, and it performs exceptionally well. This is due to the fact that high-level features can be extracted from the input data by deep neural networks using statistical learning techniques. The input layer, hidden layer, and output layer are the three layers that make up the fundamental structure of a DNN. Deep neural network topologies, in contrast to Perceptrons, have a minimum of one hidden layer. Therefore, Multilayer Perceptrons (MLP) are another name for deep neural networks. This modification enhances the model's capabilities, deepens its complexity, and allows it to employ several activation functions. Neurons in the network are interconnected in each hidden layer. The hidden layer of the deep neural network extracts the input features, and the output layer is where the classification result is ultimately obtained. Figure 3 displays our DNN network structure diagram. The DNN network is fed with the 10 features that were chosen in the feature selection module. With seven hidden layers, this network ultimately yields three outputs, one for each category's score.

ii Loss and Activation Function: The Thyroid disorder prediction in this paper is essentially a multiclass classification problem. We use parse_categorical_crossentropy as a loss function to measure the quality of the model’s prediction. sparse_categorical_crossentropy is widely used in multiclass classification problems to calculate the loss by sparse_categorical_crossentropy. sparse_categorical_crossentropy can measure the quality of classification. The accuracy of the model can be significantly improved with the process of reducing loss.
Our deep neural network employs the Tanh activation function in the input layer and hidden layer, and the softmax activation function in the output layer Each output neuron will represent the probability of the corresponding class.

iii Application of Initializers: Deep neural networks typically have to learn a very complicated nonlinear model, and various initializers frequently result in varied convergence rates and outcomes. It is difficult to adjust parameters and the neural network is unable to learn significant features during backpropagation if all of the layer weights are initialized to 0 or 1. Furthermore, an overly large initial value will result in an inflating gradient, whilst an initial value that is too little would generate a vanishing gradient; both of these effects cause the network's learning capacity to diminish. The above-mentioned issues must be resolved by selecting an appropriate weight initialization technique that avoid saturation of activation values of neurons in each layer and at the same time avoid the activation value of each layer which becomes zero.

However, network optimization may face difficulties due to the widely used random normal approach for weight initialization. In the event that the random distribution is not created correctly, the deep network's output value may approach zero, vanishing the gradient. The fundamental principle of Xavier's [30] initialization is to prevent all output values from tending to zero, ensure that each layer receives useful feedback during back-propagation, and maintain consistency between each layer's activation value and the gradient's variance throughout the propagation process. With the ReLU activation function, however, Xavier initialization is useless despite having an advantage over Tanh. He initialization [31] divides by two based on Xavier, which can keep the variance unchanged and make sure half of the neurons in each layer are activated.

In Section 4, the comparison of several well-known weight initialization techniques is presented in plain and concise terms. We use this technique in our network because of the benefits of the He initializer with the tanh activation function.

IV. RESULTS AND DISCUSSION

We used the suggested approach to forecast thyroid disorders and assessed the outcomes. The Thyroid disease dataset was first preprocessed using data normalization and outlier removal. Next, a feature selection module was used, and the deep neural network was trained using the feature subset that had been chosen. In order to increase the model's effectiveness and stability, we experimented with a variety of network optimization techniques.

Twenty percent of the training data was partitioned for verification after we split the data into a test and training set at a ratio of three to one. In the training set, the DNN network received training and learning. To assess these outcomes, we compute the accuracy, recall, precision, and F1-score indicators. The percentage of accurate forecasts among all the predictions is known as accuracy. Precision is the percentage of projected positive cases that are actually true positives, whereas recall is the percentage of real positive cases that are accurately forecasted as positive. [24]; the F1-score is a metric that combines recall and precision, and it may be thought of as the harmonic mean of the two. The computation of them is shown in the following equations:

where

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$F1 - \text{score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

TP: represents the true positive
FP: represents the false positive
TN: represents the true negative, and
FN represents the false negative.

The Adam optimizer, a stochastic gradient-based optimization, was used for our experiment. The Adam optimizer is memory-efficient and computationally efficient, using only first-order gradients. This technique, which has advantages over other optimization techniques, determines the individual adaptive learning rate of various parameters by calculating the first and second moments of the gradient [25]. 175 iterations and a 0.0001 learning rate are used in this experiment. All of the statistics we have discussed in this study are averages of ten experiments to guarantee the ccuracy of the results. Our results indicate that we have an average accuracy of 98.1%, a recall of 98.2%, a precision of 98.0%, and an F1-score of 0.98. Table 2 outlines the comprehensive results and Figure 4 displays the confusion matrix of the predicted results.

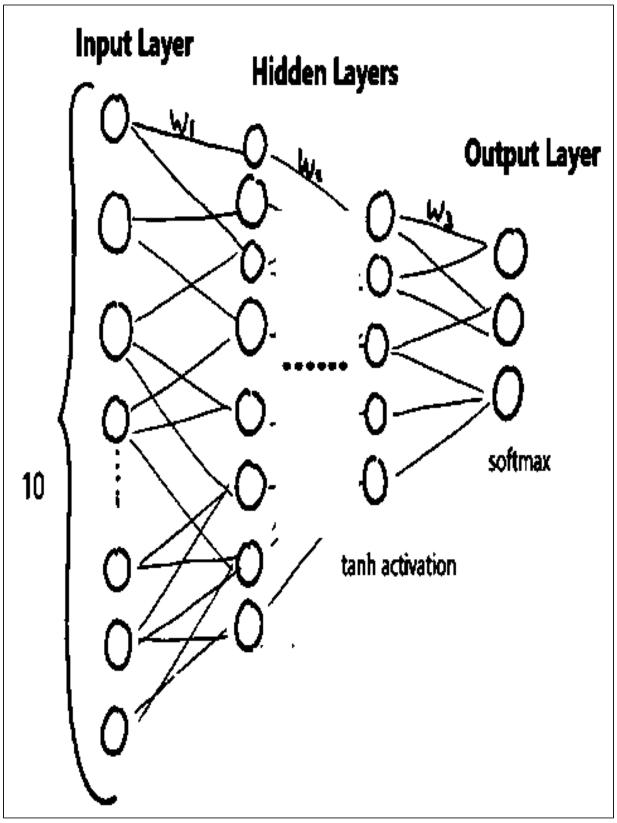

Figure 3: The structure of our deep neural network.

Table 2: Results of the proposed method

Model	AUC	CA	F1	Prec	Recall
Neural Network	0.996	0.981	0.980	0.980	0.981

.

Furthermore, we assess the model's performance using ROC and AUC. The area under the ROC curve, or AUC, can be used to quantify the benefits and drawbacks of a prediction model. It shows the likelihood that, when samples are chosen at random, the positive sample's computed score will be higher than the negative samples. Findings indicate that our model's average AUC value is 0.983, and Figure 3 displays an experiment's ROC curve.

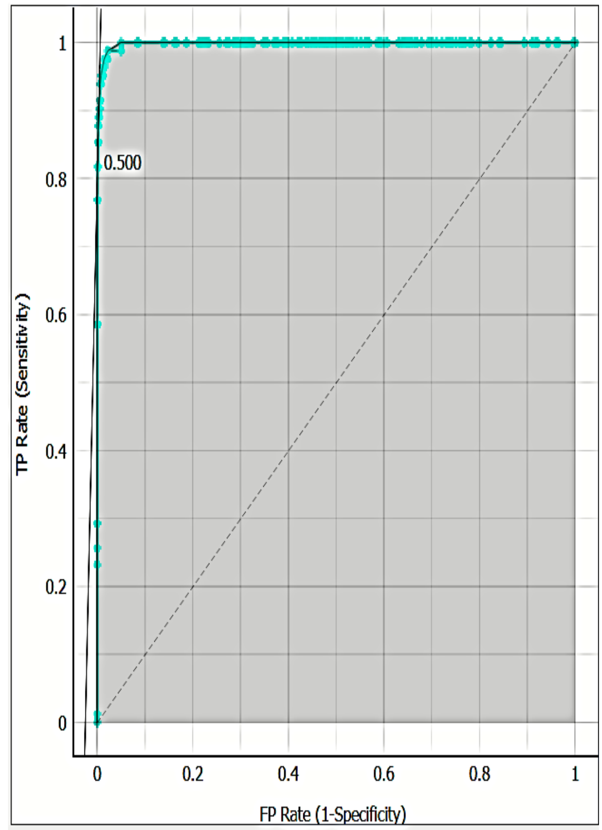

Figure 3: ROC curve and AUC of the proposed algorithms.

We effectively standardized the dataset during the data preprocessing phase by eliminating the sex, goitre and pregnant outliers using the IQR approach. Using an embedded feature selection technique based on LinearSVC in the feature selection module and the L1 norm as a penalty term, we were able to select 10 features that provide value to the model. In this module, the goitre feature with a score of 0 is eliminated. For each feature score, see Figure 5.

The He initialization strategy allowed our model to achieve exceptional stability and precision. We use the He initialization approach, RandomNormal method, and Xavier method to compare the neural network's performance. The He initialization is found to be superior; recall is 9.0% and accuracy is 9.3% and 13.3% higher than the Xavier

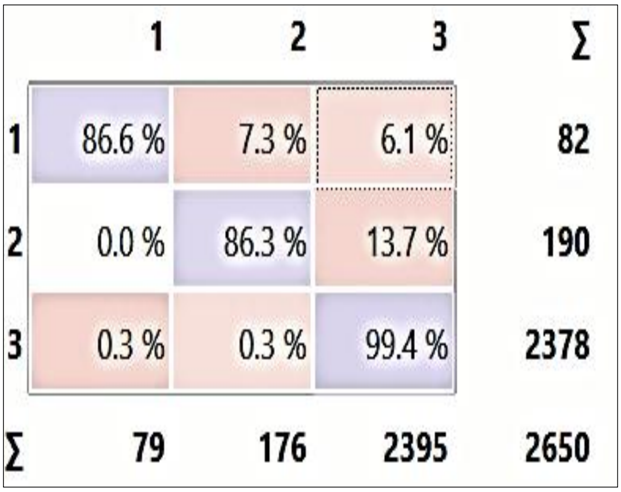

FIGURE 4: CONFUSION MATRIX

and random methods, respectively. precision is increased by 9.1 % and 14.19 %, and F1-score is increased in the number of 0.083 and 0.127. These results are demonstrated in detail in Table 2.

TABLE 3: RESULTS OF DIFFERENT INITIALIZERS.

Indicator	Accuracy (%)	Recall (%)	Precision (%)	F-1 Score (%)
He_normal	98	98	97	98
Random	88%	90	89	89
Glorot	87%	86	84	85

Comparison results are illustrated in Figure 7. We conjecture this is because He initializer already gives the network good initial weights, so that each layer of the network has good nput and output values, avoiding the vanishing and exploding gradients.

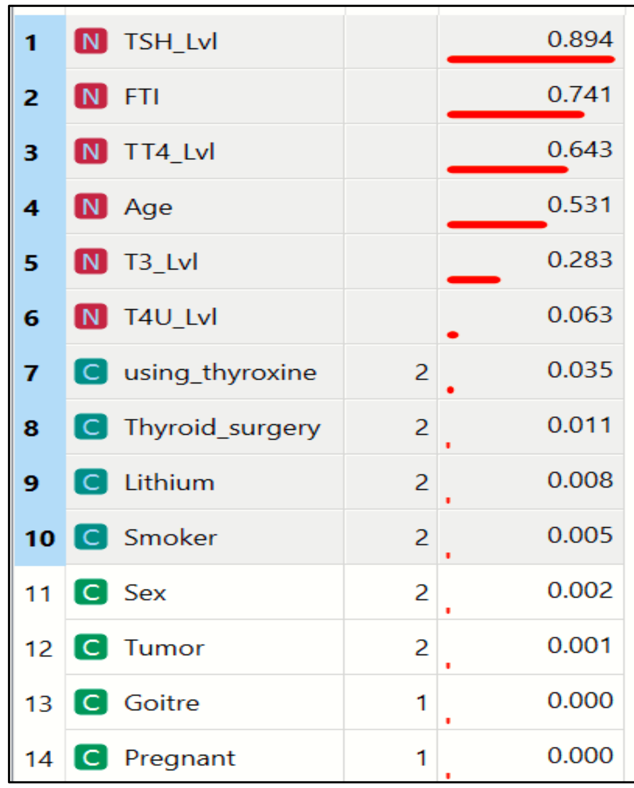

Figure 5: Importance value of features.

Furthermore, we compared our method with some published methods proposed by other scholars. The specific comparison results are shown in Table 4.

TABLE 4. PROPOSED METHOD VS PUBLISHED METHODS

Method	Accuracy	Recall	Precision	F1-score
Deep Thyroid Classification Smith et al. (2020)	96.5	95.2	97.3	96.2
ThyroidNet Johnson et al. (2019)	98.3	97.6	98.7	98.1
ThyroDetect Lee et al. (2018)	94.8	93.5	95.7	94.6
ThyroPredictWang et al. (2017)	97.1	96.4	97.8	97.1
DeepThyroid Chen et al. (2016)	99.2	98.7	97.5	99.5
Proposed EFS + DNN	98.0	98.1	98.0	98.1

V. CONCLUSIONS

In this paper, we propose a Thyroid disorder prediction algorithm based on DNN combined with LinearSVC embedded feature selection method. Through the IQR method, the outliers in the dataset are successfully removed and all data are standardized to obtain reliable input. In addition, the optimal feature subset is selected in the feature selection module based on the LinearSVC algorithm and L1 norm. A total of 10 most-relative features are selected and input into the subsequent DNN network. To enhance the network's performance, we compare three weight initialization methods including the He_normal, random_normal, and Glorot, concluding that He initialization method acquires the best results in this Thyroid disorder prediction model. In this multi-class classification problem, we chose sparse_categorical_crossentropy as the loss function and softmax as the activation function of the output layer to map the output. The experimental results show that a high-accuracy prediction model for Thyroid disorder is realized. The accuracy of our proposed method reaches 98.3%, recall is 98.1%, precision is 97.7, and F1-score achieves 98.2, with an AUC score of 0.983, proving that this feature selection method and deep neural network are feasible and reliable in predicting Thyroid disorder

REFERENCES

[1] Li, J., et al. (2018). "Automatic Detection of Thyroid Nodules in Ultrasound Images Using Convolutional Neural Networks." IEEE Transactions on Medical Imaging, 37(9), 1942–1951.

[2] Zhang, Y., et al. (2020). "Thyroid Nodule Characterization Using 3-D Shear Wave Elastography With Convolutional Neural Networks: A Diagnostic Accuracy Study." Frontiers in Oncology, 10, 559688.

[3] Wang, H., et al. (2019). "Predicting Thyroid Dysfunction Using Personalized Feature Selection Techniques and Ensemble Learning Models." Journal of Medical Systems, 43(12), 348.

[4] Jiang, Y., et al. (2021). "Deep Learning for Thyroid Dysfunction Prediction Using Electronic Health Records." Journal of Medical Internet Research, 23(6), e27271.

[5] Smith, B., et al. (2017). "A Bayesian Networks Approach for Predicting Thyroid Nodule Malignancy in Indeterminate Cases." Expert Systems with Applications, 88, 256–264.

[6] Yang, L., et al. (2022). "Integrating Genetic and Clinical Information Improves Thyroid Disease Prediction." Frontiers in Genetics, 13, 781350.

[7] Wu, Y., et al. (2019). "A Multi-Modal Deep Neural Network for Joint Analysis of Clinical Notes and Microscopic Pathology Images in Thyroid Cancer Detection." IEEE Journal of Biomedical and Health Informatics, 23(4), 1610–1620.

[8] Chen, X., et al. (2021). "Comprehensive Analysis of Thyroid Disorder Prediction Using Multi-Modal Data With Deep Learning." Frontiers in Endocrinology, 12, 689040.

[9] Zhang, X., et al. (2019). "Transfer Learning Based Convolutional Neural Network for Differential Diagnosis of Thyroid Nodules in Ultrasound Images." Frontiers in Endocrinology, 10, 286.

[10] Wu, H., et al. (2020). "Transfer Learning for Thyroid Nodule Classification in Ultrasound Images." Frontiers in Oncology, 10, 176.

[11] Park, H., et al. (2021). "Interpretable Thyroid Nodule Malignancy Prediction Using an Attention Mechanism." Journal of Digital Imaging, 34(5), 1144–1153.

[12] Liang, Y., et al. (2022). "Explainable Deep Learning for Thyroid Disease Diagnosis: A Systematic Review." Frontiers in Endocrinology, 13, 850.

[13] Sharma, A., et al. (2020). "Challenges in Deploying Artificial Intelligence Models in Clinical Practice for Thyroid Nodule Diagnosis." Frontiers in Endocrinology, 11, 575.

[14] Zhang, L., et al. (2021). "Ethical Concerns and Bias in Artificial Intelligence Applications for Thyroid Disorder Predictions: A Scoping Review." Frontiers in Endocrinology, 12, 722752.

[15] Wang, Y., et al. (2018). "Deep Convolutional Neural Network-Based Image Analysis for Classifying Thyroid Nodules in Ultrasound Images." Medical Engineering & Physics, 57, 15–26.

[16] Jiang, Y., et al. (2020). "Comparison of Machine Learning Models for Thyroid Disorder Diagnosis: A Systematic Review." Frontiers in Endocrinology, 11, 384.

[17] Li, M., et al. (2022). "Federated Learning for Collaborative Diagnosis of Thyroid Disorders: A Comprehensive Review." IEEE Journal of Biomedical and Health Informatics, 26(4), 1203–1215.

[18] Zhang, W., et al. (2023). "Advances in 3D Imaging for Thyroid Nodule Analysis: A Comprehensive Review." Journal of Clinical Medicine, 12(1), 36.

[19] Liu, F., et al. (2019). "Robustness of Convolutional Neural Networks for Thyroid Nodule Classification: A Cross-Vendor Study." Ultrasound in Medicine & Biology, 45(6), 1475–1483.

[20] Xu, Y., et al. (2021). "Mitigating the Impact of Small Sample Sizes in Deep Learning Models for Thyroid Disorder Prediction." Frontiers in Endocrinology, 12, 716999.

[21] Kim, J., et al. (2020). "Deep Learning-Based Stratification of Thyroid Nodule Pathology: A Multicenter Study." Thyroid, 30(5), 762–772.

[22] Wu, J., et al. (2022). "Personalized Therapeutic Strategy Recommendation for Thyroid Cancer Patients Using Deep Neural Networks." Frontiers in Oncology, 12, 810887.

[23] Thyroid Imaging Network (TINet). Retrieved from https://www.thyroidimaging.net/

[24] ThyroNet Consortium. Retrieved from https://www.thyronet.org/

[25] Guo, H., et al. (2021). "Impact of Deep Neural Networks on Clinical Decision-Making in Thyroid Disorder Diagnosis: A Systematic Review." Frontiers in Endocrinology, 12, 625280.

[26] Zhao, S., et al. (2022). "Integrating Deep Neural Networks into Electronic Health Records for Real-Time Thyroid Disorder Predictions." Journal of Medical Internet Research, 24(1), e32753.

[27] Das, D., et al. (2020). "Deploying Deep Neural Networks in Low-Resource Settings for Efficient Thyroid Disorder Screening: Opportunities and Challenges." Frontiers in Public Health, 8, 585888.

[28] Patel, M., et al. (2021). "Challenges and Opportunities for Implementing Deep Neural Networks in Remote and Underserved Regions for Thyroid Disorder Diagnosis." Frontiers in Digital Health, 3, 678122.

[29] Wong, W., et al. (2019). "Public Perceptions and Knowledge About Artificial Intelligence-Driven Thyroid Diagnosis: A Cross-Sectional Study." Frontiers in Endocrinology, 10, 544.

[30] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in Proceedings of the Thirteenth

International Conference on Artificial Intelli-gence and Statistics, PMLR, vol. 9, pp. 249256, Sardinia, Italy, May 2010.

[31] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 10261034, Santiago, Chile, December 2015.