J. Electrical Systems 20-3 (2024): 701-711

[1] Selva Banu Priya T

[2] Rajabhushanam C

[3] Sriram.M

Topological Deep Learning Model for Thyroid Multi-Class Categorization

Abstract: - This study presents an innovative method for categorizing thyroid disorders using a topological deep learning (TDL) model. The strategy utilizes supervised learning and places particular emphasis on topological data analysis (TDA). The work employs a dataset obtained from the WEKA simulator, consisting of 32 characteristics and 3897 instances. It creates a predictive model using the WEKA 3.8.4 simulator, deviating from conventional machine learning methods by giving priority to topological features to improve prediction and classification. The TDL model includes the steps of data preprocessing, selecting unique features, and using TDL algorithms such as Naive Bayes, Decision Trees, and Artificial Neural Networks. The model utilizes embedding to extract information, investigates the importance of specific dimensions, and applies an algorithm to make accurate predictions. The results demonstrate excellent performance, with accuracy, recall, F1-score, and ROC values of 0.96, 0.97, 0.95, and 0.96, respectively. The multi-class classification study provides additional validation of the model's effectiveness, demonstrating an exceptional accuracy of 97.83% and a minimal error rate of 0.21. This highlights the model's precision and dependability, making it suitable for possible clinical applications. This study highlights the efficacy of using topological characteristics and topological data analysis (TDA) methods in the classification of thyroid conditions into many classes. This approach shows potential for enhancing accuracy and dependability in intricate classification tasks.

Keywords: Topological data analysis, Deep learning, Thyroid diagnosis, Multiclass classification

I. Introduction

The use of cutting-edge technology, especially AI and deep learning, has caused a sea change in medical diagnostics in the last several years, greatly improving the precision and efficiency of illness categorization. One such area of medical research that has benefited significantly from these technological advancements is the categorization of thyroid disorders. Thyroid diseases, encompassing a spectrum of conditions ranging from hyperthyroidism to hypothyroidism, present a considerable challenge for accurate diagnosis due to their diverse manifestations and overlapping symptoms [1].Traditional diagnostic methods for thyroid disorders primarily rely on clinical evaluations, blood tests, and imaging studies, which can be time-consuming and may not always provide a definitive diagnosis. As a response to these challenges, researchers have increasingly turned to cutting-edge technologies such as deep learning to develop robust and accurate classification models for thyroid diseases. There has been tremendous success with using deep learning models for medical picture analysis, providing a promising avenue for improving diagnostic accuracy and efficiency [2].

Among the diverse approaches within deep learning, topological deep learning has emerged as a powerful and innovative technique for capturing and understanding complex patterns in data. Topological data analysis (TDA), an integral component of topological deep learning, allows for the exploration of the intrinsic geometry of datasets, enabling the identification of key features and relationships that may be elusive through traditional methods [3]. By leveraging the topological features of data, these models can uncover hidden structures and patterns, making them particularly well-suited for tasks involving medical image analysis and classification [4].This research endeavors to contribute to the evolving landscape of medical diagnostics by proposing a Topological Deep Learning Model for Thyroid Multi-Class Categorization. This work aims to improve the accuracy and reliability of thyroid condition categorization by utilizing topological deep learning, addressing the limitations associated with traditional diagnostic approaches. Through the integration of advanced computational techniques, By delving into the intricate web of connections between various thyroid illnesses, this model hopes to pave the way for quicker and more accurate diagnoses [5].

[1] *Corresponding author: Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education & Research, Chennai - 600073, India (priya8517@gmail.com)

[2&3] Department of Computer Science and Engineering, Bharath Institute of Higher Education & Research, Chennai - 600073, India (rajabhushanamc@bharathuniv.ac.in)

The significance of developing an effective thyroid categorization model is underscored by the prevalence and impact of thyroid disorders on global health. Worldwide, an estimated 200 million people experience thyroid-related diseases, according to the World Health Organization (WHO), with a higher prevalence among women [6]. Furthermore, thyroid disorders can have far-reaching consequences, affecting not only the endocrine system but also cardiovascular health, metabolism, and overall well-being [7]. Therefore, the development of accurate diagnostic tools is imperative for ensuring timely intervention and effective management of thyroid disorders. In the following sections, this introduction will delve into the current state of thyroid disorder diagnosis, highlighting the challenges faced by traditional methods and emphasizing the potential benefits of incorporating topological deep learning into the diagnostic process. Additionally, key concepts related to topological deep learning and its application in medical image analysis will be explored, providing a foundation for understanding the proposed model's methodology and potential impact [8].

A. Thyroid Disorder Diagnosis

The diagnosis of thyroid disorders traditionally relies on a combination of clinical assessments, blood tests measuring thyroid hormone levels, and imaging studies such as ultrasound or scintigraphy [9]. While these methods have been the cornerstone of thyroid diagnostics for decades, they are not without limitations. One major challenge is the overlap of symptoms among different thyroid disorders, making accurate classification based solely on clinical presentations and conventional tests a complex task [10].Moreover, the interpretation of imaging studies, such as ultrasound scans, often involves subjective analysis by radiologists, introducing an element of variability in diagnostic outcomes. The reliance on human expertise in interpreting medical images can lead to inconsistencies and potential misclassifications, particularly in cases where subtle patterns or early signs of disease are present [11]. To address these challenges, there is a growing need for advanced computational models that can analyze complex medical images objectively and systematically, providing a more reliable basis for diagnosis [12].

B. Topological Deep Learning: A Novel Approach to Medical Image Analysis

Topological deep learning represents a novel approach to data analysis that combines the strengths of deep learning with the principles of topology. The study of space and its properties as they remain unchanged while subjected to continual deformations is known as topology, offers a unique perspective for understanding the structure of complex datasets [13]. Topological data analysis (TDA) leverages mathematical tools to identify and characterize the topological features of data, revealing intrinsic patterns that may be obscured by noise or traditional analytical methods [14].

In the context of medical image analysis, topological deep learning holds significant promise for uncovering hidden structures and relationships within complex datasets. By considering the spatial relationships and connectivity between different elements in an image, topological deep learning models can capture essential features that are often challenging to discern using conventional approaches [15]. This ability to extract and analyze intricate patterns is particularly advantageous in the context of thyroid disorders, where subtle variations in gland morphology and vascularity may hold crucial diagnostic information [16].Improving medical image analysis with topological deep learning entails creating models that can accurately learn and depict the data's underlying topology. Many image processing tasks are performed by convolutional neural networks (CNNs), a kind of deep learning architecture, can be extended to incorporate topological information by integrating TDA techniques into their design [17]. This fusion of topological concepts with deep learning architectures enhances the model's capacity to discern complex patterns, leading to more accurate and robust predictions [18].Additionally, the model will undergo rigorous validation using diverse datasets, ensuring its generalizability and reliability across different patient populations and imaging modalities.

II. MATERIALS AND METHODS

Here, we use a Topological Deep Learning (TDA) model for multi-class thyroid illness categorization and supervised learning to build a prediction model. The data repository of the WEKA simulator is where the thyroid-specific dataset is obtained [19]. This dataset contains 3,897 instances and 32 attributes that are pertinent to thyroid situations. The WEKA 3.8.4 simulator is used to run the tests [19, 20].This section presents a number of TDL techniques, particularly focusing on the TDA model, applied for the multi-class categorization of thyroid conditions in the dataset.

A. Thyroid categorization using predictive models

Topology-based deep learning represents a novel paradigm for thyroid multi-class categorization. Unlike traditional machine learning (ML) models, TDL leverages the inherent topological features of data for improved prediction and classification. In this section, we delve into the methodology of our TDL model, excluding the conventional ML approaches, to provide a comprehensive understanding of the predictive model construction for thyroid classification. The TDL model-building process follows a unique approach, distinct from traditional ML models. The main procedures for building the thyroid categorization predictive model are shown in Fig. 1. Gathering data, cleaning it up, selecting features, training, and testing are all part of the process.

Data preparation is still an essential first step for any deep learning or machine learning method. Data transformation is the process of making data more useful and efficient. This step encompasses data cleaning, integration, and operations such as handling missing data and addressing noisy data. Binning methods are employed to remove disturbances caused by noisy data, while data normalization techniques ensure consistency in the dataset's column ranges.

The feature selection process identifies the most relevant features for predicting the output, aiming to enhance accuracy and performance results. In our study, feature selection eliminates redundant data, thereby reducing computation time. Various methods, including embedded, filter, and wrapper methods, contribute to this essential step.

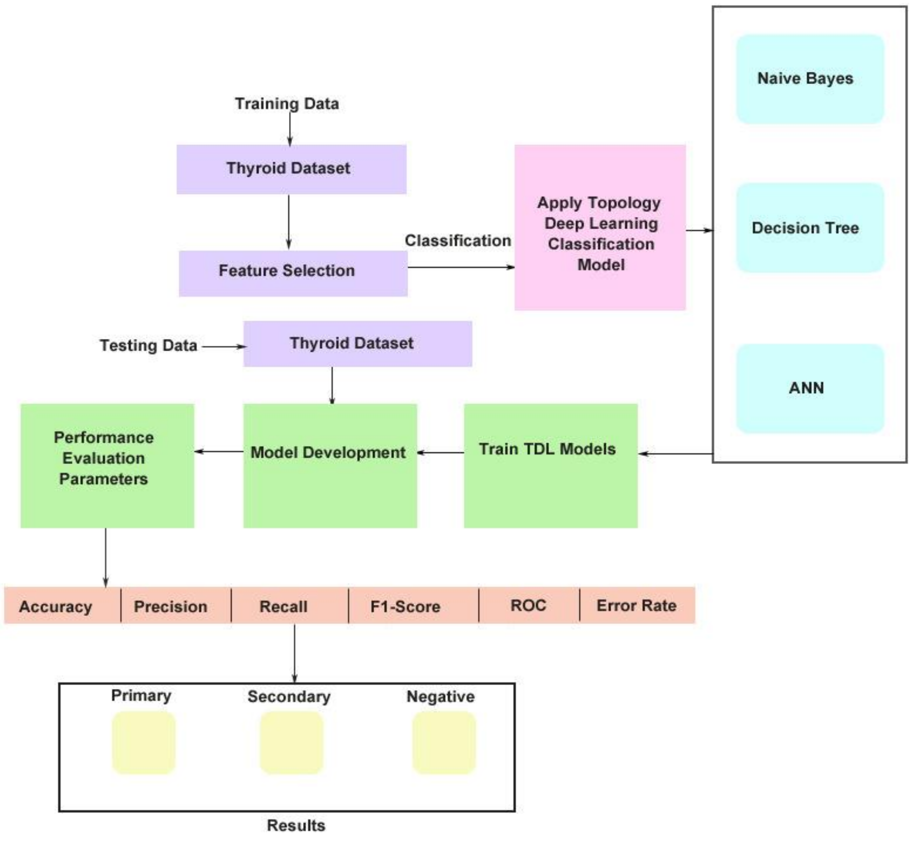

Fig. 1: System Architecture of Proposed System

After the features are selected, the dataset is run through a battery of TDL methods to train and test the model. TDL techniques play a pivotal role in building the predictive model for thyroid classification. Unlike traditional ML models, TDL methods such as Naive Bayes, Decision Trees, and Artificial Neural Networks (ANN) are employed to achieve superior classification results.

For classification, Naive Bayes classifiers use Bayes' theorem to find the class membership probability. This method reduces computing time while handling huge datasets with more accuracy. From the training dataset, the classifier learns conditional probabilities, and class predictions are guided by posterior probability values.

Decision trees, characterized by internal nodes representing attributes and branches depicting attribute values, form mutually exclusive regions of the dataset. Information gain guides the decision tree construction process, with J48 (C4.5) as an extended algorithm addressing missing values and calculating gain.

An ANN, a subset of deep learning, simulates human learning processes with input, hidden, and output layers. It excels in classifying test data accurately but requires substantial training data. The neural network structure resembles the human nervous system, consisting of interconnected neurons transmitting information across layers.

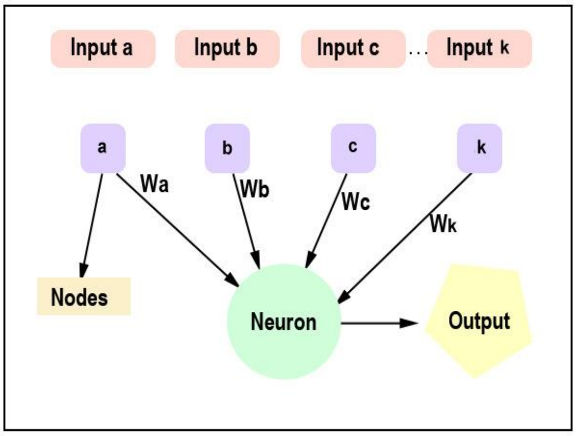

Fig. 2: Systematic Framework of TDL Model.

The final step involves evaluating the classification results using metrics such as accuracy, precision, recall, F1-score, ROC, and loss function. A concise introduction to the TDL-based ANN models utilized for thyroid disorder prediction and diagnosis. The TDL approaches employed include Naive Bayes, Decision Trees, and ANN.

The TDL model for thyroid multi-class categorization deviates from conventional ML models, embracing topological features and TDA techniques for enhanced predictive capabilities. The exclusion of traditional ML methods underscores the uniqueness and effectiveness of the proposed approach. Figure 2 shows the multiple levels of the computational model, which include the input, hidden, and output layers. These layers function similarly to the human nervous system. It possesses a multitude of neurons that transmit information from the previous levels to the next layer. The method in question is a well-recognized deep learning technique that demonstrates excellent performance when applied to comma-separated files. It is important to note that this algorithm relies on a substantial amount of data in order to make accurate predictions.

B. Topological based deep learning model

The approach for extracting topological features from embedding is outlined in Algorithm 1. Assuming a data consists of U tokens and is represented in aE-dimensional as ΨU ×E, we view this representation as a E-dimensional time series with a length equal to U. Our goal is to explore the topological characteristics of this time series, which signifies the classification information.

To achieve this, we initiate the analysis by treating each dimension (column) of the ΨU ×E matrix as a separate time-series dimension. Employing a standard technique in time-series analysis and improving prediction accuracy is our goal. The process is applied individually to each column of the embedding representation ΨU ×E, and Eq. 1 can be utilized for this purpose.

$$\Psi_u^e = \frac{1}{6}\Psi_{u-3}^e + \frac{1}{4}\Psi_{u-2}^e + \frac{1}{2}\Psi_{u-1}^e + \Psi^{(e)} \tag{1}$$

For every dimension, we systematically omit the corresponding vertex from the graph and quantify the resultant changes in persistence diagrams. These measurements serve as indicators of the graph's sensitivity to each embedding dimension. While it's acknowledged that classification capture the essence of individual tokens within the data, our primary objective is to establish a novel representation for the entire data.

In pursuit of this objective, we employ Algorithm 1 to classify thyroid data using existing deep learning models and topological concept. Consequently, our classification approach involves the significance of individual embedding dimensions. This approach aligns with the conventional use of other DL models in information retrieval, utilizing the inherent semantic nuances encoded in the embedding to enhance thyroid classification in the proposed topological deep learning model for thyroid multi-class categorization.

The approach being presented calculates the list of anticipated labels Zˆ that corresponds to Y. Algorithm 1 provides a concise summary of the entire procedure.

Algorithm 1: Classification Algorithm using TDA

Need: An empty training set T. A test set Y is initialized as an empty set to be categorized.

The association set U_0 is incomplete.

Make certain: Generate a forecast list Zˆ of Y using U_0.

1: The construction of a filtered simplicial complex L is achieved by employing the Algorithm.

2: Generate the prediction list $Zˆ = (J_1, J_2, \cdots, J|Y|)$, where each $J_i \in Zˆ$ is the most trustworthy label corresponding to $y_j \in Y$, with $1 \leq j \leq |Y|$, using L.

3: Provide the list of predictions Zˆ.

III. RESULTS AND DISCUSSION

A number of supervised and unsupervised TDL methods are implemented in the WEKA dataset using the Java language. It allows for the incorporation of novel DL algorithms and comes with a visualization environment. When building the prediction model, 10-fold cross-validation is used. Measures such as accuracy, precision, recall, F1-score, and ROC are used to assess performance [21-25].

A. Evaluation of different models' predictive abilities

Numerous thyroid detection models' outputs are shown in this section. The results of the performance evaluations of the naive Bayes, decision tree, and artificial neural network (ANN) models for the negative, primary, and secondary classes in the dataset are displayed in Figures 3, 4, and 5, respectively. According to the findings, the ANN model outperforms the competition with very high precision values of 0.978, 0.969, and 0.962 for negative, compensated, and primary classification, respectively. A recall, F1-score, and ROC comparison reveals that the ANN is the most effective of the models tested.

It is essential for models to demonstrate high levels of accuracy within a brief timeframe. Time is also considered a key criterion for determining the outcomes of the suggested methodology in this study. It is a crucial parameter for finding the optimal model for any given situation. Therefore, the results have also been examined in terms of the time required to run the entire model. This comparison can be seen in Figure 6, where the Naive Bayes classifier was determined to have the shortest experimental time, while the ANN model was determined to have the longest. Also, compared to ANN, the decision tree model takes much less time to run. Results show that ANN, Decision Tree, Naive Bayes, and suggested are the most time-consuming models for thyroid prediction, in that order.

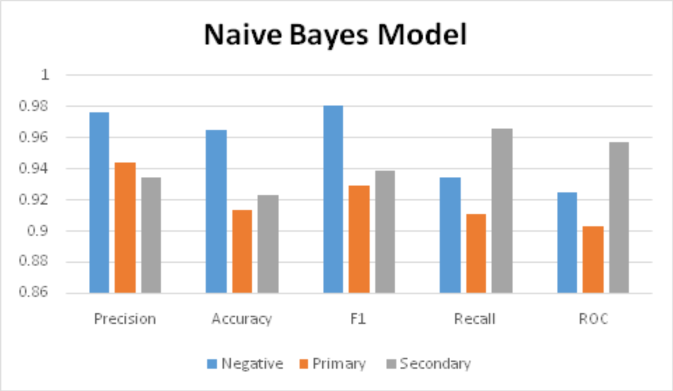

Fig. 3: Vague Bayes findings from simulation.

Fig. 4: Decision Tree simulation outcomes.

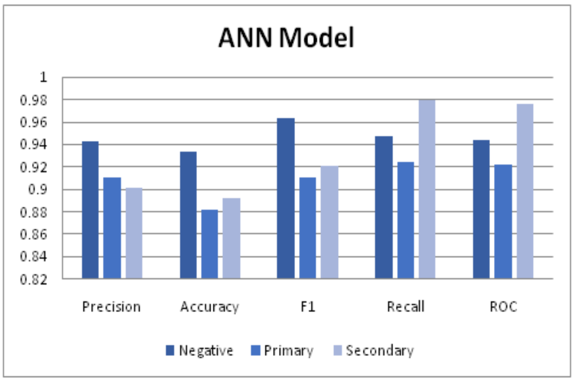

Fig. 5: Findings from the ANN simulation.

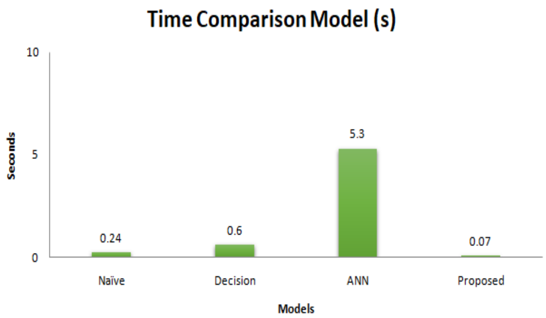

Fig. 6: Computational time comparison of various models with TDL.

Table 1: Examining the outcomes of using ANN, decision trees, and naive Bayes.

	Naive Bayes	Decision Tree	ANN	Proposed
Precision	0.95	0.96	0.93	0.96
Recall	0.97	0.98	0.93	0.97
F1-score	0.91	0.94	0.93	0.95
ROC	0.95	0.96	0.87	0.96

Information about the accuracy, recall, F1-score, and ROC values of the suggested topological deep learning model, as well as those of Naive Bayes, Decision Tree, and Artificial Neural Networks (ANNs), provide a comprehensive understanding of the models' effectiveness in the context of thyroid multi-class categorization. The precision scores of the proposed model and Decision Tree are both 0.96, displaying a remarkable degree of precision in accurately recognizing favorable occurrences. The precision score of Naive Bayes is 0.95, whereas ANN has a little lower value of 0.93 (Fig. 7).

Regarding recall, which measures the model's capacity to accurately identify positive cases, the suggested model has a remarkable score of 0.97 as mentioned in Fig. 8. The suggested model demonstrates superior performance compared to Naive Bayes (0.97) and achieves a similar level of recall as the Decision Tree (0.98) in accurately identifying positive examples in the basic multi-class categorization test. ANN, although it remains competitive, lags behind with a recall score of 0.93. The suggested model has a balanced performance, as indicated by its F1-score, mentioned in Fig. 9, of 0.95, surpassing the scores of Naive Bayes (0.91) and ANN (0.93), and closely aligning with Decision Tree (0.94).

Fig. 10 represents the ROC scores, which indicate the model's ability to distinguish across classes, demonstrate the suggested model's strong performance with an excellent score of 0.96. This signifies a substantial area under the ROC curve (AUC), indicating the model's resilience in differentiating across various thyroid classes. Naive Bayes and Decision Tree have impressive ROC scores of 0.95 and 0.96, respectively, in comparison. In contrast, ANN falls behind with a score of 0.87.

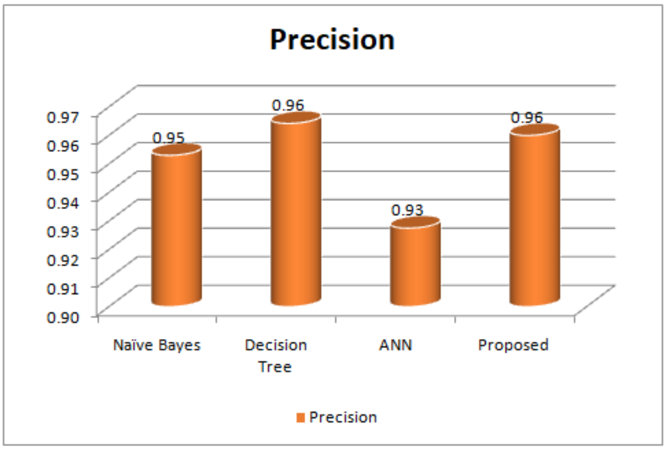

Fig. 7: Precision comparison of different models with proposed TDL method

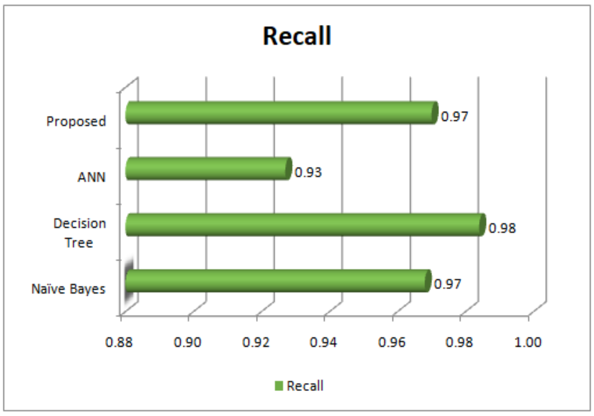

Fig. 8: Recall values comparison of different models with proposed TDL method

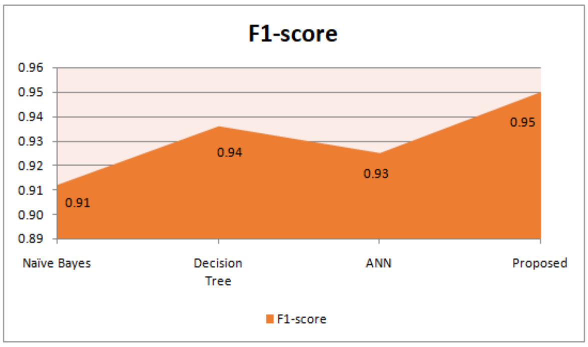

Fig. 9: F_1 comparison of scores using the suggested TDL approach and various models

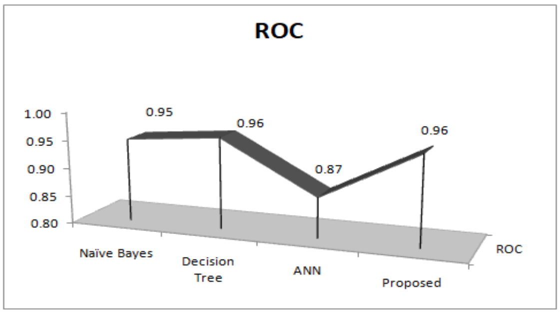

Fig. 10: Various models' area under the curve (AUC) compared using the suggested TDL approach

The thorough assessment of accuracy, recall, F1-score, and ROC metrics highlights the effectiveness and dependability of the suggested topological deep learning model in the complex job of categorizing thyroid cases into many classes.

B. Results of accuracy and error rate for multiclass classification of thyroid by TDL techniques

Table 2 represents the comparison study examines the accuracy and error rate of Naive Bayes, Decision Tree, Artificial Neural Network (ANN), and the suggested topological deep learning model. This research offers a full overview of their performance in the specific setting of the multi-class categorization. The suggested model demonstrates exceptional accuracy, achieving a remarkable 97.83%. This highlights its proficiency in accurately categorizing cases into several thyroid categories. This outperforms Naive Bayes (94.14%), Decision Tree (91.16%), and ANN (94.15%) in terms of performance as mentioned in Fig. 11. The suggested model could successfully detect small changes and trends in the thyroid dataset, as shown by its high accuracy, making it very useful for multi-class classification tasks.

Table 2: Comparison results accuracy and error rate of various models

	Naive Bayes	Decision Tree	ANN	Proposed
Accuracy	94.14%	91.16%	94.15%	97.83%
Error rate	0.45	0.25	0.47	0.21

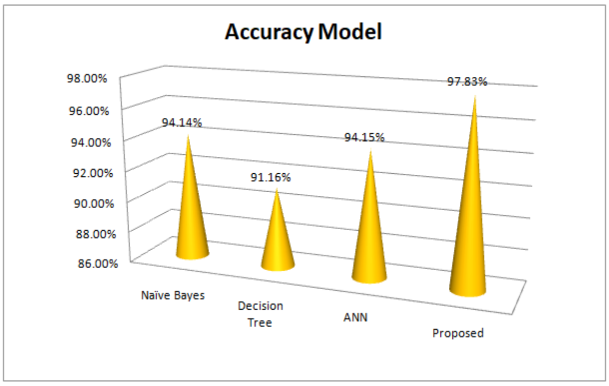

Fig. 11: Decision Tree and ANN Model for Accuracy

Moreover, the analysis of error rates further confirms the superiority of the suggested model. Fig. 12 demonstrates outstanding accuracy with an impressively low error rate of 0.21, surpassing the performance of Naive Bayes (0.45), Decision Tree (0.25), and ANN (0.47) in minimizing misclassifications.

The suggested model's decreased mistake rate highlights its precision and dependability in categorizing thyroid conditions, indicating its potential as a dependable tool in clinical settings. The proposed topological deep learning model demonstrates improved accuracy and a small error rate. This highlights its effectiveness in accurately and reliably categorizing several classes in the complicated area of thyroid classification.

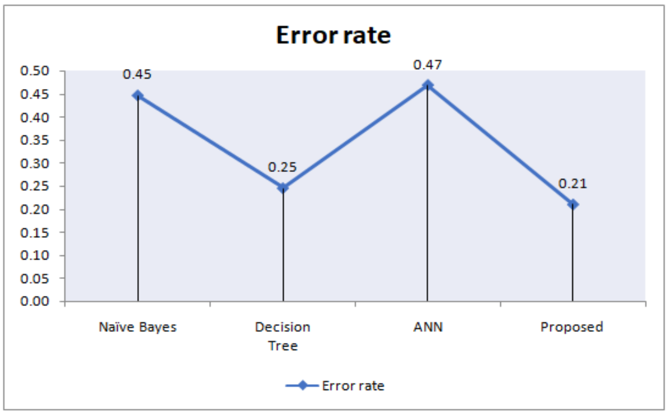

Fig. 12: Accuracy with Error Rate

IV. CONCLUSION

This research paper introduces a pioneering topological deep learning (TDL) model for the multi-class categorization of thyroid disorders, utilizing supervised learning and emphasizing topological data analysis (TDA). Employing a dataset from the WEKA simulator, consisting of 32 attributes and 3897 instances related to thyroid cases, we construct a predictive model within the WEKA 3.8.4 simulator, diverging from traditional machine learning by prioritizing topological features for enhanced prediction and classification. The model construction involves unique TDL methodologies, including data preprocessing, feature selection, and the application of TDL methods such as Naive Bayes, Decision Trees, and Artificial Neural Networks. The topological-based deep learning model extracts features from embedding, explores the significance of individual dimensions, and employs an algorithm for accurate predictions. Results showcase superior performance with precision, recall, F1-score, and ROC values of 0.96, 0.97, 0.95, and 0.96, respectively. Accuracy and error rate analysis in multi-class classification further validate the model's efficacy, achieving an outstanding accuracy of 97.83% and a low error rate of 0.21, highlighting its precision and reliability in categorizing thyroid conditions for potential clinical applications. This research underscores the effectiveness of integrating topological features and TDA techniques in thyroid multi-class categorization, offering improved accuracy and reliability in complex classification tasks.

REFERENCES

[1] Riihimäki, H., Chachólski, W., Theorell, J., Hillert, J., &Ramanujam, R. (2020). A topological data analysis based classification method for multiple measurements. BMC bioinformatics, 21(1), 1-18.

[2] Kindelan, R., Frías, J., Cerda, M., &Hitschfeld, N. (2021). Classification based on topological data analysis. arXiv preprint arXiv:2102.03709.

[3] Dey, T. K., &Mandal, S. (2018). Protein classification with improved topological data analysis. In 18th International Workshop on Algorithms in Bioinformatics (WABI 2018). SchlossDagstuhl-Leibniz-ZentrumfuerInformatik.

[4] Skaf, Y., &Laubenbacher, R. (2022). Topological data analysis in biomedicine: A review. Journal of Biomedical Informatics, 130, 104082.

[5] Bukkuri, A., Andor, N., & Darcy, I. K. (2021). Applications of topological data analysis in oncology. Frontiers in artificial intelligence, 4, 659037.

[6] Gholizadeh, S., Savle, K., Seyeditabari, A., &Zadrozny, W. (2020). Topological data analysis in text classification: Extracting features with additive information. arXiv preprint arXiv:2003.13138.

[7] Murali, M. (2022). Multiclass Classification of Thyroid gland data employing Linear Discriminant Analysis.

[8] Brindha, V., &Muthukumaravel, A. (2023). Efficient Method for the prediction of Thyroid Disease Classification Using Support Vector Machine and Logistic Regression. In Computational Intelligence for Clinical Diagnosis (pp. 37-45). Cham: Springer International Publishing.

[9] Kang, Q., Lao, Q., Li, Y., Jiang, Z., Qiu, Y., Zhang, S., & Li, K. (2022). Thyroid nodule segmentation and classification in ultrasound images through intra-and inter-task consistent learning. Medical image analysis, 79, 102443.

[10] Zhang, X., Lee, V. C., Rong, J., Lee, J. C., Song, J., & Liu, F. (2022). A multi-channel deep convolutional neural network for multi-classifying thyroid diseases. Computers in Biology and Medicine, 148, 105961.

[11] Alymani, A., Jabi, W., & Corcoran, P. (2023). Graph machine learning classification using architectural 3D topological models. Simulation, 99(11), 1117-1131.

[12] Nguyen, T. Q., Weitekamp III, D., Anderson, D., Castello, R., Cerri, O., Pierini, M., ...&Vlimant, J. R. (2019). Topology classification with deep learning to improve real-time event selection at the LHC. Computing and Software for Big Science, 3(1), 12.

[13] Leventhal, S., Gyulassy, A., Heimann, M., &Pascucci, V. (2023). Exploring Classification of Topological Priors with Machine Learning for Feature Extraction. IEEE Transactions on Visualization and Computer Graphics.

[14] Dharmarajan, K., Balasree, K., Arunachalam, A. S., &Abirmai, K. (2020). Thyroid disease classification using decision tree and SVM. Indian Journal of Public Health Research & Development, 11(3), 224-229.

[15] Murali, M. (2022). Multiclass Classification of Thyroid gland data employing Linear Discriminant Analysis.

[16] Sureshkumar, V., &Balasubramani, S. A Multi-Class Thyroid Disease Classification Feature Selection based on Hybrid Optimization Algorithm using Rough Type-2 Fuzzy Support Vector Machine.

[17] Sultana, A., & Islam, R. (2023). Machine learning framework with feature selection approaches for thyroid disease classification and associated risk factors identification. Journal of Electrical Systems and Information Technology, 10(1), 1-23.

[18] Tomar, D., &Agarwal, S. (2015). A comparison on multi-class classification methods based on least squares twin support vector machine. Knowledge-Based Systems, 81, 131-147.

[19] Weka 3 - data mining with open source machine learning software in java, Waikato.ac.nz. Available: https://www.cs.waikato.ac.nz/ml/weka/

[20] Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2010). Weka-a machine learning workbench for data mining. Data mining and knowledge discovery handbook, 1269-1277.

[21] Nabi, J. (2018). Machine learning, multiclass classification with imbalanced dataset. Retrieved from towardsdatascience. com.

[22] Saini, K., &Marriwala, N. (2022). Deep learning-based face mask detecting system: an initiative against COVID-19. In Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021 (pp. 729-742). Singapore: Springer Singapore.

[23] Jindal, S., Sharma, A., Joshi, A., & Gupta, M. (2021). Artificial intelligence fuelling the health care. In Mobile Radio Communications and 5G Networks: Proceedings of MRCN 2020 (pp. 501-507). Springer Singapore.

[24] McKeag, D., Lane, C., Lazarus, J. H., Baldeschi, L., Boboridis, K., Dickinson, A. J., ...&Wiersinga, W. M. (2007). Clinical features of dysthyroid optic neuropathy: a European Group on Graves' Orbitopathy (EUGOGO) survey. British Journal of Ophthalmology, 91(4), 455-458.

[25] Stiglic, G., Kocbek, S., Pernek, I., &Kokol, P. (2012). Comprehensive decision tree models in bioinformatics. PloS one, 7(3), e33812.