International Journal for Modern Trends in Science and Technology
Volume 10, Issue 04, pages 348-353.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue04.html
DOI: https://doi.org/10.46501/IJMTST1004053

Machine Learning-Based Classification of Thyroid Disorders: Advancing Diagnostic Capabilities

M.Raja Kumar[1] , Ungarala Sai Praneetha[2], Chodavarapu S V.V.Sai Kumar[2], Mulleti Varsha[2], Soothi Sriram[2], Addanki Snehitha Naga Sai Lalitha[2]

[1]Associate Professor, Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.
[2]Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.

To Cite this Article
M.Raja Kumar , Ungarala Sai Praneetha, Chodavarapu S V.V.Sai Kumar, Mulleti Varsha, Soothi Sriram, Addanki Snehitha Naga Sai Lalitha, Machine Learning-Based Classification of Thyroid Disorders: Advancing Diagnostic Capabilities, International Journal for Modern Trends in Science and Technology, 2024, 10(04), pages. 348-353. https://doi.org/10.46501/IJMTST1004053

Article Info
Received: 06 April 2024; Accepted: 18 April 2024; Published: 26 April 2024.

ABSTRACT

Particularly in the healthcare industry, machine learning algorithms and data mining approaches are essential for handling the enormous volumes of data and information that exist today. Depending on the severity of the condition, thyroid abnormalities can now result in a wide range of health issues. By categorizing thyroid diseases using machine learning techniques, the current effort seeks to improve the procurement of thyroid disorders. This study aims to categorize this illness into eight groups: the condition, a lack of thyroid normal, and five other types. Using advanced machine learning techniques, customized loss functions, and feature extraction, this study offers a Multi-layer Neural Network model that classifies the thyroid illness data set.The proposed technique takes use of eight class classifiers to improve sickness classification. The recommended model's performance is evaluated using parametric models such as SVM and a data set of 9172 samples with 30 attributes. Several model versions based on the TensorFlow functional API were employed.

***Keywords:** thyroid disorder, machine learning, neuralnetwork, tensor flow, auto encoders, PCA, SVM*

1. INTRODUCTION

Among endocrine illnesses, thyroid dysfunction is currently the most poorly understood and diagnosed group. Thyroid issues are the second most prevalent endocrine disorders worldwide [1], only behind diabetes, according to the WHO. About 2% of people have hyperthyroidism, whereas 1% have hypothyroidism. Thyroid issues have also affected millions of people in India; an estimated 42 million Indians are thought to be affected by thyroid diseases [5], based on various studies and research. Experts say that preventing disease progression and possibly death requires early sickness identification, diagnosis, and treatment. For a variety of illnesses, early detection and differential diagnosis increase the likelihood of successful therapy.Clinical diagnosis is frequently seen as a tough task, even after multiple tests. The thyroid gland, located near the base of the throat, resembles a butterfly. The hormone complex contains levothyroxine (T4) and triiodothyronine (T3), both of which are required by the brain to regulate body temperature, blood pressure, and heart rate. Thyroid disease, like iodine deficiency, is one of the most frequent ailments in the world. However, it is possible that other factors are at play. As an endocrine gland, the thyroid secretes hormones into the bloodstream. It is situated at the body's frontal midline.Hormones produced by the thyroid gland aid in digestion, bodily equilibrium, and moisture retention. Triiodothyronine (T3), thyroid hormone (T4), and thyroid stimulating hormone (TSH) are thyroid gland therapies [3]. Thyroid function is evaluated with TSH administration. Thyroid cancer, goiter, iodine deficiency, Hashimoto's thyroiditis, and hyperthyroidism are the five most prevalent thyroid conditions in India. An abortion might result from this sickness if treatment is delayed, which is quite concerning for expectant mothers who have trouble controlling their thyroxin levels. Early disease identification has never been simpler thanks to machine learning, which has made it easier to forecast sickness using test cases and test data.Deep learning algorithms [4] are among the most effective approaches for tackling a variety of connected complex problems. However, after reviewing the literature, it is discovered that classifiers are primarily designed for binary classification, which uses the numbers 0 and 1 to represent two classes, such as no thyroid and thyroid, or three class classification, which uses the numbers 0 and 1 to represent three classes, hypothyroid and hyperthyroid. The review reveals that the bulk of machine learning models being created are non-parametric models, and that research is mostly focused on three to four distinct forms of thyroid disorders. Following comprehensive research, a deep neural network model built using unique feature engineering methodologies was presented.to create an eight-class classifier, three of which are anti-thyroid treatment, miscellaneous, and no thyroid, with the remaining five being the ones already supplied. The first section of this work is an introduction; the second is a review of the literature that addresses current advances in this field of research and gap analysis; and the third section is about the architecture of the proposed categorization model. The fourth section discusses the performance of the recommended SVM-based classifier, while the last portion provides the study's conclusion and prospective future applications.

2. LITERATURESURVEY

Chandel, Kushboo, [6] and associates first employed unsupervised grouping techniques, including K-nearest neighbor, to create a multiclass classification model that could distinguish between different types of thyroid disorders. They employed a number of supervised learning techniques, including support vector machines, KNN, and the Naïve Bayes Classifier. Using a Rapid miner tool, the study trial was conducted, and the results showed that K-Nearest Neighbour has the highest accuracy (93.44 percent) for identifying thyroid ailment. Rasitha Banu [7], G.: Several data mining techniques are applied using the WEKA tool to characterize hypothyroidism. They used the data mining classification algorithms J48 and DecisionStudget. The data repository at the University of California provided the data utilized in this investigation.Their research revealed that J48 outperforms the second one, with an accuracy of 99.58 percent, surpassing the ensemble method referred to as decision stump. SVM, Decision Trees, and Naïve Bayes are examples of native classification techniques that Umar Sidiq et al. [8] employed to detect different thyroid illnesses. They reported an accuracy of 98.89 using Aanconda3 version 5.2.0. Sindhya and Mrs. K claimed [9] to have used J48 to achieve a 99 percent accuracy rate while classifying thyroid illnesses. Additionally, they said that it ran in 0.2 seconds, which is faster than Random Forest models.GoksuAkgul, Ali AknCelik, ZelihaErgulAydin, and ZehraKamisliOzturk [11] developed a data mining-based technique to better accurately diagnose hypothyroidism. Before building the hypothyroidism diagnostic models, many sampling strategies were utilized to eliminate the imbalanced distribution from the data. These techniques included Logistic Regression, K Nearest Neighbour, and Support Vector Machine. The logistic regression classifier performed best after being trained on the data set using oversampling approaches, with a success rate of 97.8%, an F-Score of 82.26 percent, a zone under the curve of 93.2%, and a Matthews coefficient of correlation of 81.8 percent.Vijaya Kumar et al. develop a machine learning model that accurately and precisely predicts diabetes in a patient at an early stage using the Random Forest technique. Random

Forest algorithms are widely utilized for regression and classification in ensemble learning systems. The outcomes demonstrate that the prediction method can accurately, quickly, and most significantly, forecast the onset of diabetes. The best results for diabetes prediction are obtained with the suggested approach. The breast cancer research by V. Chaurasia, S. Pal, and B. B. Tiwari [12] inspired them to employ state-of-the-art techniques and tools.Breast cancer is the most common cancer in women, accounting for more than any other kind. In light of recent breakthroughs (683 breast cancer cases), they developed prediction models to improve the model for breast cancer survivability utilizing a large dataset and three well-known data mining methods (Nave Bayes, RBF Network, and J48). They used 10-fold cross-validation techniques to evaluate the unbiased estimates of the three prediction models and compare their accuracy. This study employed two criteriabenign and malignant cancer casesto evaluate three breast cancer survival prediction models. Numerous studies on thyroid classification have been undertaken, and the literature reveals that these studies employ a range of data mining techniques to create reliable classifiers.Four classification approachesNaive Bayes, Decision Tree, MLP, and RBF Networksare proposed in the research for thyroid disorders, including hyperthyroidism and hypothyroidism. For thyroid data, these conditions are referred to as maximal classification models. The accuracy rate with which various thyroid classification models use their classification methodologies.

3. SYSTEM ANALYSIS

A. EXISTING SYSTEM

The current "Thyroid Disorder Classification using Machine Learning" technique uses a multi-layer neural network model to categorize thyroid ailments into eight classes, including hyperthyroidism, hypothyroidism, normal, and five more. To boost classification accuracy, the system employs feature extraction, customized loss functions, and advanced machine learning techniques. To improve performance, a number of model variants based on the Tensor Flow functional API and innovative feature engineering methodologies are applied. Using a dataset of 9172 samples with 30 features, the proposed model is compared against parametric models such as SVM. The goal is to enhance the diagnosis and categorization of thyroid problems by leveraging cutting-edge machine learning approaches to address the obstacles given by health data.

DISADVANTAGES OF THE EXISTING SYSTEM

1. **Limited Generalization**: When applied to different patient populations or datasets that are not part of the original training set, the current system might not perform as well as it might. When applied to different healthcare environments or separate demographic groupings, this limitation may lead to subpar performance.
2. **Data Imbalance**: Biased model training may emerge from imbalances in the dataset's distribution of thyroid issue categories. The model's ability to appropriately identify these less prevalent disorders may suffer if particular classes are underrepresented.
3. **Interpretability Issues**: The interpretability of the results may be limited by the complexity of the multi-layer neural network model. In healthcare applications, it is essential to comprehend the rationale behind certain predictions; nevertheless, the model's opaque nature might impede its acceptance and inspire mistrust among medical professionals.
4. **Dependency on Feature Engineering**: The system is based on unconventional feature engineering approaches, which may require a significant amount of computing power and may not deliver the optimal feature selection. Relying too heavily on bespoke characteristics may limit the model's ability to adapt to new healthcare data.
5. **Scalability Challenges**: The scalability of the proposed methodology can be problematic when working with larger datasets or in real-time clinical settings. Over time, the system's processing requirements might make it less effective and timely in handling more and more health data.

B. PROPOSED SYSTEM

By combining a number of substantial modifications, the proposed technique aims to address the inadequacies of the present framework for identifying thyroid illnesses. To begin, a more comprehensive strategy to data collection will be implemented to decrease data imbalances and improve the model's ability to generalize over a wide variety of patient demographics. Second, by using explainable AI approaches, interpretability will be prioritized, ensuring that medical staff can understand and trust the model's predictions. Third, the proposed system will look into more complicated deep learning architectures than the Multi-layer Neural Network, taking into consideration cutting-edge models that may yield superior results. Furthermore, research will be conducted on automated feature extraction approaches to reduce the requirement for human feature engineering and better flexibility with changing healthcare datasets. Ultimately, the model's computational efficiency will be increased to better address scalability and make it more appropriate for real-time clinical applications. In general, the proposed method aims to provide a more precise, comprehensible, and expandable solution for thyroid condition classification, hence fostering the expansion of machine learning applications in the medical field.

ADVANTAGES OF THE PROPOSED SYSTEM

1. **Improved Classification Accuracy**: The proposed method enhances classification accuracy for thyroid diseases by using cutting-edge deep learning architectures and powerful machine learning algorithms. This innovation ensures a more reliable diagnostic tool for medical practitioners.
2. **Enhanced Generalization**: The proposed system addresses data imbalances and enhances its ability to generalize across different patient groups through the use of a strong data collection approach. This guarantees the accuracy and efficacy of the model in a variety of healthcare settings and demographic groups.
3. **Interpretability and Trustworthiness**: Explainable AI systems, which prioritize interpretability, provide medical practitioners with insight into the model's decision-making process. This transparency builds acceptance and trust among practitioners, which is critical for the successful use of machine learning in healthcare.
4. **Adaptability to Evolving Data**: By eliminating the requirement for human feature engineering, automated feature extraction techniques enable the model to gradually adjust to changes in healthcare data. Because of its flexibility, the system remains efficient and useful even as medical databases expand.
5. **Scalability for Real-time Applications**: The proposed system is scalable due to improvement of processing efficiency, making it suitable for real-time clinical applications. This advantage is critical for providing timely and sensitive decision support in medical settings, which improves patient care and outcomes.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

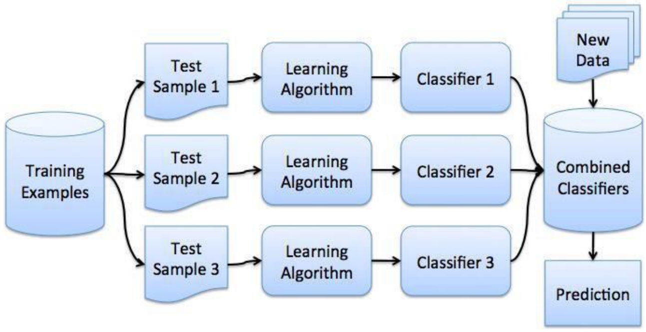

Fig 1. Methodology followed for proposed model

5. SYSTEM IMPLEMENTATION

MODULES

The following is an outline of the modules that make up the proposed "Thyroid Disorder Classification using Machine Learning" system.

Data Collection and Preprocessing: Obtaining a varied dataset with samples related to thyroid disorders.

phases in the preprocessing process that deal with class imbalances, missing data, and normalizing.

Feature Engineering: Use of machine-learning-based feature extraction approaches to find pertinent trends from the raw data; exploration of innovative feature engineering methodologies to better the portrayal of characteristics linked with thyroid problems[11].

Model Development: Many deep learning frameworks have been created and applied, including neural network configuration other than the multi-layer neural network[12].

Using advanced machine learning techniques to enhance the model's classification skills.

Customized Loss Functions: Customized loss functions designed to address the unique problems and peculiarities of thyroid disease categorization.

Explainable AI Techniques: Incorporate explainable AI technologies to increase model interpretability, allowing medical practitioners to better understand and accept the system's predictions.

Performance Evaluation: Conducting rigorous assessments utilizing measures like as accuracy, precision, recall, and F1 score to analyze how the model performed on the dataset.

Comparison with Parametric Models:A comparative examination using standard parametric models like Support Vector Machines (SVM) to test the performance of the proposed deep learning-based technique[10].

Scalability Optimization: Implement ways to enhance the model's computational efficiency, guaranteeing scalability for real-time applications in clinical situations.

Validation and Testing: The model was rigorously validated using cross-validation techniques to demonstrate its resilience and generalizability to previously unexplored data. Test the system on several test datasets to confirm its real-world applicability.

Documentation and Deployment: The created system is comprehensively documented, including the model architecture, training procedure, and assessment findings. The finished model will be deployed for practical usage in healthcare facilities, including considerations for usability and integration into current systems.

6. RESULTS AND DISCUSSION

The analysis of principal components selects characteristics that are maximum scattered and orthogonal to each other. In our suggested model, PCA outperformed all other feature extraction strategies (see table). Our suggested model achieved 92.36 percentage accuracy for eight class categorizations. The findings are presented in Table 3. The graphs in figures 5 and 6 depict the losses that happened when training the model using Auto The encoders and PCA as feature extraction strategies, respectively. The Autoencoder obtained a minimal loss of 0.4478, whereas PCA achieved a loss of 0.2475.

tabular form below

TABLE III: RESULTS OF PROPOSED DEEP LEARNING MODEL

S. No.	Framework/ Algorithm	Feature Engineering Techniques	Accuracy
1	TensorFLow Functional API	Feature selection using chi-square estimation techniques	88.76
2	TensorFLow Functional API	Feature selection using Autoencoders	85
3	TensorFLow Functional API	Feature Extraction using PCA	**92.36**
4	Support Vector Machines	Feature Selection with RBF as Kernel	75.53
5	Support Vector Machines	Feature Selection with Polynomial as Kernel	78.31
6	Decision Tree Classifier	Feature Selection	87.67

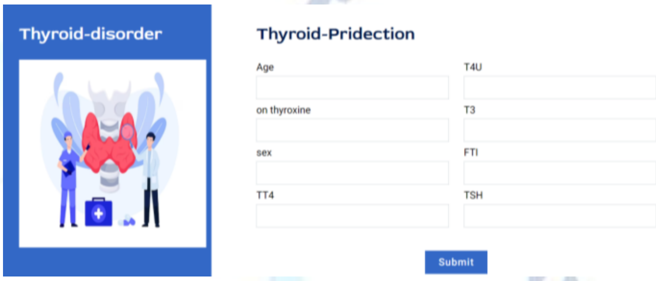

Fig 2. Providing input data for prediction

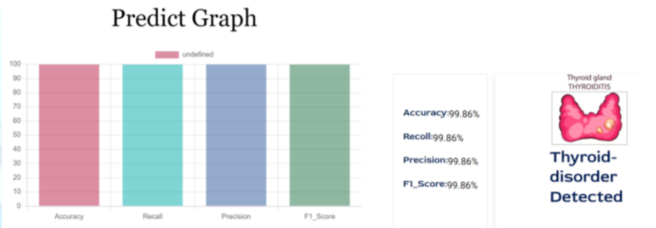

Fig 3. Results

7. CONCLUSION AD FUTURE WORK

SVM generated an accuracy of about 92 percent for four class classification, whereas MLP produced an accuracy of 96 percent for three classes. SVM has certain limitations; it does not perform well with multiclass classification issues or overlapping data sets. The MLP model, without any feature engineering, achieved an accuracy of 86 percent. This proposed study was built utilizing a multi-layer neural network model that utilized advanced feature engineering approaches to enhance the performance of deep classifiers. This research proposes a Multi-layer Neural Network model that employs Customized Loss Functions and advanced ML and Feature Engineering approaches to attain an accuracy of 92.36 percent for eight categories classifications. In future results can be further analyzed using Ensemble Methods.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Azar, a.T, Hassanien, A.E. and Kim, "Expert system based neural fuzzy rules thyroid disease diagnosis Computer Science, Artificial Intelligence", arXiv:1403.0522, Pp. 1-12,2012

[2] Chandani Patel Chavez, Maria del Mar Morales Hernandez, Jesse Kresak and Whitney W. Woodmansee," Evaluation of multi nodular goitre and primary hyperparathyroidism leads to a diagnosis of AL amyloidosis Division of Endocrinology", Diabetes and Metabolism, University of Florida, Gainesville, FL, USA.

[3] Bashar Bazkke, Joli Osman, Mohammad Shahrour, Mohammad Ziadeh, Aya Haji Mohamad, Mohamed ImadEddinMouhandes&AmmarNiazi, A pregnant women with history of hashimoto's thyroiditis diagnosed with Kikuchi-Fujimoto disease: the first case report, University of Florida, Gainesville, FL, USA.

[4] Khalid Salman et al., "Thyroid Disease Classification using Machine Learning Algorithm hm", 2021 J. Phys.: Conf. Ser. 1963 012140

[5] Ambika Gopalakrishnan Unnikrishnan, Usha V. Menon , "Thyroid disorders in India: An epidemiological perspective", Indian J EndocrinolMetab. 2011 Jul; 15(Suppl2): S78–S81. doi: 10.4103/2230-8210.83329

[6] Chandel, Kushboo, et al. "A comparative study on thyroid disease detection using K-nearest neighbour and classification techniques". CSI transactions on ICT ICT 4.2-4 (2016): 313-319.

[7] Banu, G. Rasita et al. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease" International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.

[8] Umar Sidiq et al., Diagnosis of various thyroid ailments using data mining classification technique". Int Journal SCI RES Coput SCI InfTecnol 5 (2019) : 131-6

[9] Aswad, Salma Abdullah, and EmrullahSonuç. "Classification of VPN Network Traffic Flow Using time Related Features on Apache Spark." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020

[10] Manjula Devarakonda Venkata, Sumalatha Lingamgunta," Triple-Modality Breast Cancer Diagnosis and Analysis in middle aged women by Logistic Regression", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-4, February 2019

[11] Manjula Devarakonda Venkata, Sumalatha Lingamgunta, "Statistical and Texture Features based Breast Cancer Classification from Multi image Modalities using Support Vector Machines with Gaussian Radial Basis Function", International Journal of Control and Automation Vol. 13, No.1, (2020), pp. 384 - 396

[12] Shaik Mohammad Rafee, Manjula Devarakonda Venkata," An Ensemble Neuro Fuzzy Algorithm For Breast Cancer Detection And Classification", Ictact Journal On Soft Computing, January 2024, Volume: 14, Issue: 03, DOI: 10.21917/ijsc.2024.0460

[13] Manjula Devarakonda Venkata, Sumalatha Lingamgunta Health Care Automation in Compliance to Industry 4.0 Standards :A Case Study of Liver disease Prediction", Journal of Scientific & Industrial Research Vol. 82, February 2023, pp. 263-268 DOI: 10.56042/jsir.v82i2.70215