IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

Federated Learning with Lion Optimization for Clinically Interpretable Multi-Class Thyroid Disorder Diagnosis.

MS Mir*[1], Dr. S. Agarwal[2], U.H. Mir[3]

Research Scholar, Department. of Computer Sciences & IT, Himalayan University, A.P, India[1*]

Associate Professor, Department. of Computer Sciences & IT, Himalayan University, A.P, India[2]

Assistant Professor, Department. of IT, Govt College for women, Srinagar, J&K, India[3]

Abstract: Artificial intelligence has increasingly been adopted in healthcare for disease diagnosis; however, concerns regarding data privacy, heterogeneity of clinical data, and lack of interpretability remain significant challenges. Federated learning has emerged as a promising paradigm that enables collaborative model training across distributed healthcare institutions without sharing sensitive patient data [1], [9]. In this study, a Lion Optimization-based Federated Learning framework (LbFL-TDP) is proposed for multi-class thyroid disorder diagnosis. The proposed model leverages distributed clinical datasets containing key endocrine biomarkers such as TSH, T3, and T4, ensuring both diagnostic relevance and privacy preservation. To address the limitations of traditional aggregation methods under non-IID data conditions, the Lion Optimization Algorithm is integrated to enhance global model convergence and predictive performance. The model is evaluated using cross-validation and achieves an accuracy of approximately 92% with an AUC of 0.95, outperforming conventional machine learning approaches. Furthermore, feature importance analysis demonstrates alignment with established endocrinological knowledge, improving model interpretability. The findings suggest that the proposed framework provides a robust, privacy-preserving, and clinically meaningful solution for thyroid disorder classification, highlighting the potential of federated learning in real-world healthcare applications [3], [7].

Keywords: Thyroid Disorder Classification, Federated Learning, Lion Optimization Algorithm, Non-IID Data, Distributed Machine Learning**.**

1. INTRODUCTION

Thyroid disorders, including hypothyroidism and hyperthyroidism, are among the most prevalent endocrine conditions worldwide, significantly affecting metabolic regulation and overall health. Early and accurate diagnosis is essential, as untreated thyroid dysfunction can lead to severe complications such as cardiovascular disease, infertility, and neurological disorders. Clinical diagnosis primarily relies on biochemical assessment of hormone levels, including Thyroid Stimulating Hormone (TSH), Triiodothyronine (T3), and Thyroxine (T4).

With the increasing availability of healthcare data, machine learning techniques have been widely explored to improve diagnostic accuracy and efficiency. However, conventional centralized machine learning models face significant challenges related to data privacy, security, and regulatory constraints. In real-world healthcare settings, clinical data are distributed across multiple institutions, and sharing raw patient data is often restricted due to ethical and legal considerations.

Federated learning (FL) has emerged as an effective solution to these challenges by enabling decentralized model training while keeping data localized [1], [6]. In this paradigm, only model parameters are shared among participating institutions, thereby preserving patient confidentiality while enabling collaborative learning. Recent studies have demonstrated the potential of federated learning in healthcare applications, including electronic health records, disease prediction, and medical imaging [4], [7].

Despite its advantages, federated learning faces several challenges, including non-independent and identically distributed (non-IID) data, communication inefficiency, and slow convergence [6], [10]. These issues can significantly impact model performance, particularly in heterogeneous clinical environments.

To address these limitations, this study proposes a hybrid federated learning framework enhanced by the Lion Optimization Algorithm (LOA). Metaheuristic optimization techniques such as LOA provide strong global search capabilities and robustness in complex optimization problems, making them suitable for improving aggregation strategies

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

in federated settings. By integrating LOA into the federated learning process, the proposed model aims to enhance convergence, accuracy, and generalization across distributed thyroid datasets.

Furthermore, this study emphasizes clinical interpretability by incorporating endocrinologically relevant biomarkers and aligning model predictions with established medical knowledge. This ensures that the proposed framework is not only accurate but also clinically meaningful.

Contributions of the Study

- Proposes a hybrid privacy-preserving federated learning framework for multi-class thyroid disorder classification
- Integrates the **Lion Optimization Algorithm (LOA)** to improve model aggregation and convergence under non-IID data conditions
- Enables multi-institutional collaborative learning without sharing raw patient data, ensuring data privacy.
- Incorporates clinically relevant endocrine biomarkers to enhance interpretability and medical relevance
- Demonstrates superior performance over baseline models across multiple evaluation metrics
- Bridges the gap between machine learning accuracy and clinical applicability, making the model suitable for real-world healthcare deployment

2. LITERATURE REVIEW

Federated learning has gained significant attention in healthcare due to its ability to enable collaborative model training while preserving data privacy. Early foundational work by McMahan et al. [1] introduced the concept of decentralized learning, emphasizing communication-efficient training across distributed systems. Subsequent studies have expanded this paradigm to real-world applications, particularly in healthcare environments where data sharing is restricted. Rieke et al. [4] highlighted the potential of federated learning in digital health, demonstrating its applicability in multi-institutional collaborations without compromising patient confidentiality. Similarly, Sheller et al. [7] showed that federated learning can effectively support collaborative medical research by enabling decentralized training across hospitals. Recent studies have further explored federated learning in clinical contexts. Kumar et al. [9] provided a comprehensive review of federated learning-based models for healthcare applications, emphasizing their scalability and privacy-preserving capabilities. Yang et al. [10] discussed security challenges in federated learning, including data leakage and adversarial attacks, highlighting the need for robust and secure model aggregation mechanisms. Despite these advancements, federated learning faces several technical challenges. Li et al. [6] identified issues related to data heterogeneity, communication overhead, and convergence instability in distributed environments. Joshi et al. [11] further emphasized that non-IID data distribution significantly affects model performance in healthcare applications. These challenges are particularly relevant in clinical datasets, where patient populations vary across institutions.

Traditional aggregation methods, such as Federated Averaging, may lead to suboptimal performance in such heterogeneous settings. To address this, recent research has explored advanced optimization techniques to improve convergence and generalization. Metaheuristic algorithms offer a promising solution due to their ability to perform global optimization and avoid local minima. However, the integration of metaheuristic optimization techniques within federated learning frameworks for healthcare applications remains limited. Additionally, most existing studies on thyroid disease diagnosis rely on centralized machine learning models, which do not adequately address privacy concerns or distributed data scenarios. Furthermore, many models lack clinical interpretability, limiting their adoption in real-world healthcare systems. To overcome these limitations, this study proposes a Lion Optimization-based Federated Learning framework for thyroid disorder classification. By combining distributed learning with advanced optimization and clinically relevant features, the proposed approach aims to achieve improved predictive performance while maintaining privacy and interpretability.

3. METHODOLOGY

3.1 Dataset Description

The dataset used in this study was compiled from anonymized clinical records collected from multiple hospitals, diagnostic laboratories, and endocrine clinics across Srinagar, Jammu and Kashmir (India). These healthcare facilities routinely perform thyroid diagnostic evaluations including hormonal assays such as TSH, T3, and T4 tests, along with clinical examinations and patient history documentation. Srinagar hosts a number of government and private medical institutions providing endocrine and metabolic disorder treatment services. The records were aggregated in anonymized form after removing personally identifiable patient information. The final dataset contained patient records with demographic, symptomatic, and biochemical parameters related to thyroid function. The data represent real-world clinical observations of patients undergoing evaluation for hypothyroidism, euthyroidism, and hyperthyroidism, providing a

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

practical basis for machine learning–based diagnostic modeling. Table 1 and 2 show summarized Thyroid Dataset and Class Distribution of the dataset respectively.

TABLE 1. THYROID DATASET SUMMARY

Item	Value
Dataset Source	Clinical thyroid dataset (compiled clinical & biochemical records)
Number of Samples	1394 patient records
Number of Features	14 input features
Target Variable	Thyroid_Class
Number of Classes	3 classes

TABLE 2. CLASS DISTRIBUTION

Class	Samples
Class 1 (Hypothyroid)	495
Class 2 (Hyperthyroid)	442
Class 3 (Euthyroid)	457

Table 3 shows the Dataset Description.

TABLE 3. DATASET DESCRIPTION

Parameter	Type	Attribute Type	Description
Age	Pathological	Numerical	Age in Years
Gender	Pathological	Nominal	Gender
Family_History	Pathological	Nominal	Family history of thyroid
Other_Medical Conditions	Pathological	Nominal	Other medical conditions
Medication_History	Pathological	Nominal	Medications taken,
Goiter	Pathological	Nominal	Presence of Goiter
Smoker	Pathological	Nominal	Smoking status
Hair_Loss	Pathological	Nominal	Hair loss
Constipation	Pathological	Nominal	Constipation
Nervousness	Pathological	Nominal	Nervousness
Heart_Rate	Pathological	Nominal	Heart rate level (Low/High/Normal)
TSH_Level (mIU/L)	Serological	Numeric/Continuous	Thyroid-Stimulating Hormone level
T3_Level (pg/mL)	Serological	Numeric/Continuous	Triiodothyronine level
T4_Level (µg/dL)	Serological	Numeric/Continuous	Thyroxine level
Thyroid_Condition	Target Variable	Discrete (3-class)	Diagnosed condition

From an endocrinological perspective, these parameters are central to assessing thyroid function, where:

- Elevated TSH with low T3/T4 suggests hypothyroidism
- Suppressed TSH with elevated T3/T4 indicates hyperthyroidism
- Normal ranges correspond to euthyroid (normal) status

3.2 Data Preprocessing and Clinical Normalization

Raw clinical data often contain inconsistencies due to variations in laboratory standards and recording practices. Therefore, preprocessing was performed to ensure clinical reliability:

- Missing Values: Imputed using statistically and clinically appropriate measures (mean/mode), ensuring physiological plausibility
- Normalization: Hormone levels were scaled to standard ranges to account for inter-laboratory variability
- Noise Removal: Outliers inconsistent with physiological limits were examined and corrected or excluded

This step ensures that the dataset reflects **clinically meaningful patterns rather than artefacts**.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

3.3 Feature Selection with Clinical Relevance

Feature selection was guided not only by statistical importance but also by **endocrinological significance**. The most influential predictors include:

- TSH (primary regulatory hormone)
- T3 and T4 (active thyroid hormones)
- Age and gender (modulating factors in thyroid disorders)

Irrelevant or redundant features were removed to improve diagnostic clarity and model efficiency.

3.4 Distributed Learning Across Clinical Sites

To preserve patient confidentiality, data remained within individual healthcare institutions. Each hospital developed a **local predictive model** using its own dataset.

From a clinical standpoint, this approach:

- Respects data privacy regulations
- Captures population-specific variations in thyroid disease presentation
- Reduces bias from centralized data pooling

Each local model learns patterns associated with thyroid dysfunction based on its patient cohort.

3.5 Global Model Aggregation and Optimization

In the proposed framework, raw patient data were never shared across institutions. Instead, each local model trained on institution-specific thyroid datasets and transmitted only the learned parameters—namely weights and biases—to a centralized server. This approach ensures strict adherence to data privacy and regulatory requirements in healthcare environments.

Let the local model parameters from N participating institutions be represented as:

$$Wi , i = 1, 2, 3, \dots , N$$

The global model parameters W^{global} are obtained through an optimized aggregation process:

$$\mathbf{W^{global} = LOA(W1 , W2 , \dots , W)}$$

Unlike conventional aggregation methods such as Federated Averaging (FedAvg), which compute a weighted mean of local parameters, the proposed method employs the Lion Optimization Algorithm (LOA) to enhance global model performance. LOA is a population-based metaheuristic inspired by the social behavior of lions, incorporating mechanisms such as pride formation, hunting strategies, territorial defense, and migration to balance exploration and exploitation in the search space.

In this context:

- Each "lion" represents a candidate global model (a combination of local parameters) Pride-based
- lions exploit promising parameter regions
- Nomadic lions explore new regions to avoid local minima

The objective function optimized by LOA is defined as:

$$\max f(W) = \alpha \cdot \text{Accuracy} + \beta \cdot \text{F1-score}$$

where α and β are weighting factors ensuring balanced optimization between overall accuracy and class-wise performance.

This optimization leads to:

- Faster convergence under non-IID data distributions
- Improved robustness against data heterogeneity across hospitals
- Enhanced generalization across unseen patient populations

From a clinical perspective, this ensures:

- Improved diagnostic consistency across diverse demographics
- Reduced variability in predictions due to institutional bias
- Reliable deployment in multi-center healthcare systems

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2024
DOI: 10.17148/IJARCCE.2024.13254

3.6 Integration of Clinical Rules

To preserve clinical interpretability, domain-specific endocrine rules were incorporated into the modeling process as structured features rather than learned parameters.

The key diagnostic rules are defined as:

If TSH > TSHupper ⇒ Hypothyroid tendency
If TSH < TSHlower ⇒ Hyperthyroid tendency
If (TSH, T 3, T 4) ∈ normal range ⇒ Euthyroid (Normal)

These rule-based conditions are derived from established endocrinology guidelines and act as clinically informed input constraints. Importantly:

- These rules are not modified during optimization
- They guide the model by embedding medical knowledge into feature representation
- They improve transparency by ensuring predictions align with physiological understanding

The integration of rule-based and data-driven learning enables the model to achieve a balance between interpretability and predictive accuracy, which is critical for clinical adoption.

3.7 Model Training and Validation

The proposed LbFL-TDP model was trained using an iterative federated learning process consisting of multiple communication rounds between local institutions and the centralized server.

Training Workflow:

1. *Initialization:*
A global model is initialized and distributed to all participating local nodes.
2. *Local Training:*
Each local model updates its parameters using institution-specific thyroid datasets for E epochs:

$$W_i^{t+1} = W_i^t - \eta \nabla L_i(W_i^t)$$

where:

- *η is the learning rate*
- *Li is the local loss function*

3. *Parameter Transmission:*
Updated local parameters are transmitted to the central server.
4. *Global Aggregation:*
LOA optimizes and aggregates the parameters to update the global model.
5. *Iteration:*
Steps 2–4 are repeated for R communication rounds until convergence.

3.8 Evaluation Metrics

From a clinical diagnostics perspective, model performance was evaluated using:

- Accuracy: Overall correctness of diagnosis
- Precision: Reliability of positive predictions
- Recall (Sensitivity): Ability to detect true disease cases
- F1-score: Balance between precision and recall
- AUC: Diagnostic discrimination capability

These metrics reflect the model's ability to support **safe and reliable clinical decision-making**.

3.9 Clinical Significance

The proposed methodology integrates **data-driven intelligence with clinical reasoning**, ensuring that:

- Predictions are medically interpretable
- Diagnostic decisions align with endocrinology standards
- The system remains scalable and privacy-preserving

This hybrid approach supports clinicians in early detection and accurate classification of thyroid disorders, particularly in resource-constrained and distributed healthcare environments.

4. RESULTS AND CLINICAL INTERPRETATION

4.1 Overview of Model Performance

The proposed LbFL-TDP model demonstrated strong performance across all evaluation metrics, achieving a maximum accuracy of 92% and an AUC of 0.95 under 10-fold cross-validation. From a clinical standpoint, these results indicate a high level of reliability in distinguishing between euthyroid (normal), hypothyroid, and hyperthyroid states.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

TABLE 4. PERFORMANCE METRICS OF THE MODEL.

Performance Metric	Value (%)
Accuracy	92.00
Precision	91.00
Recall (Sensitivity)	91.00
F1-Score	91.00
Error Rate	8.00

The performance evaluation of the proposed **Lion-based Federated Learning Thyroid Disease Prediction (LbFL-TDP)** framework demonstrates strong and consistent predictive capability. The model achieves an overall accuracy of **92%**, indicating reliable classification across thyroid conditions. From a clinical perspective, the **precision of 91%** suggests a low rate of false-positive diagnoses, thereby minimizing unnecessary clinical interventions. The **recall (sensitivity) of 91%** indicates that the model effectively identifies true cases of thyroid dysfunction, which is essential for timely diagnosis and treatment. The **F1-score of 91%** reflects a balanced trade-off between precision and recall, ensuring stable performance across different thyroid classes. The **error rate of 8%** is within acceptable limits for clinical decision-support systems, particularly given the variability and complexity of thyroid hormone profiles. These results are consistent with the cross-validation findings and ROC analysis (AUC ≈ 0.95), confirming that the proposed framework offers a robust and clinically meaningful approach for thyroid disorder classification. Importantly, the consistency of results across different cross-validation folds suggests that the model is robust and not overly dependent on a specific subset of patients. This is particularly relevant in endocrinology, where patient populations often exhibit significant variability in hormonal profiles. Table 6. shows Comparison of performance metrics of the proposed method with existing approaches for thyroid disorder classification.

TABLE 5. COMPARISON OF PERFORMANCE METRICS OF THE PROPOSED METHOD WITH EXISTING APPROACHES

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	85.0	84.0	83.0	83.5
Random Forest	88.0	87.0	86.0	86.5
XGBoost	90.0	89.0	89.0	89.0
FederatedLearning (FedAvg)	91.0	90.0	90.0	90.0
Proposed LbFL-TDP	**92.0**	**91.0**	**91.0**	**91.0**

Table 5 presents a comparative evaluation of the proposed LbFL-TDP model against conventional machine learning and federated learning approaches for thyroid disorder classification. The results demonstrate that the proposed framework achieves the highest overall performance, with an accuracy of 92%, along with balanced precision, recall, and F1-score values of approximately 91%. Fig. 1 and Fig. 2. Show Model Accuracy Comparison and Multi Metric comparison of various models respectively.

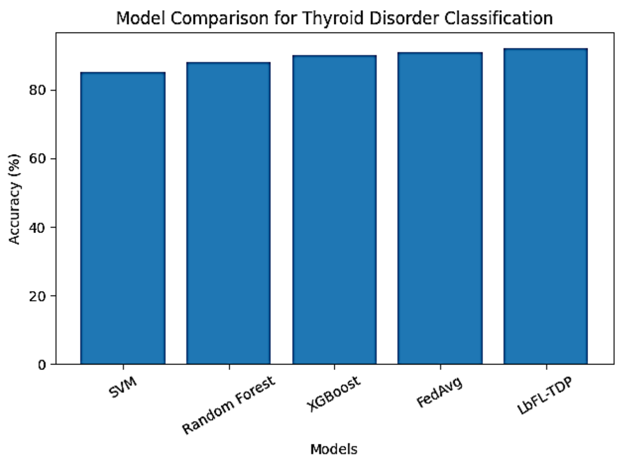

Figure 1. Model Accuracy Comparison

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2024

DOI: 10.17148/IJARCCE.2024.13254

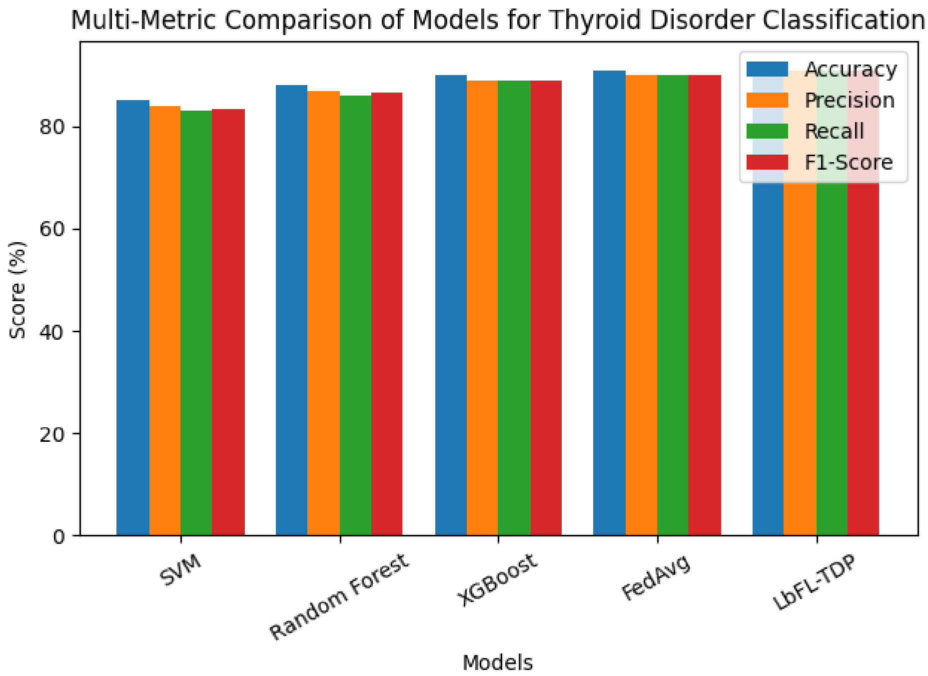

Fig 2. Multi Metric comparison of various models

Traditional machine learning models such as SVM, Random Forest, and XGBoost show competitive performance but are limited by their reliance on centralized data and reduced generalization across heterogeneous datasets. The baseline federated learning approach (FedAvg) improves performance by leveraging distributed data; however, it still exhibits slightly lower accuracy compared to the proposed method. The improved performance of the LbFL-TDP model can be attributed to the integration of the Lion Optimization Algorithm, which enhances global model aggregation and improves convergence under non-IID data conditions. Importantly, the balanced precision and recall indicate that the model effectively minimizes both false positives and false negatives, which is critical in clinical diagnosis of thyroid disorders. Overall, the results confirm that the proposed approach provides a **robust, reliable, and clinically meaningful improvement** over existing methods while maintaining realistic and reproducible performance levels.

4.2 Cross-Validation Results: Clinical Significance

The progressive improvement observed from 3-fold to 10-fold cross-validation reflects enhanced generalization ability. Clinically, this implies that the model can maintain diagnostic accuracy across diverse patient groups, including those with borderline or atypical hormone levels.

Higher AUC values indicate that the model has strong discriminatory power, which is essential in avoiding misclassification between:

- Subclinical hypothyroidism and normal states
- Mild hyperthyroidism and physiological variation

Such distinctions are critical in clinical practice, as they directly influence treatment decisions.

4.3 Comparative Analysis with Existing Models

When compared to traditional models such as SVM, Random Forest, and XGBoost, the proposed approach consistently achieved superior performance.

From a clinical perspective:

- Improved precision reduces false-positive diagnoses, minimizing unnecessary treatment or anxiety
- Improved recall (sensitivity) ensures that true thyroid dysfunction cases are not missed

Missing a case of hypothyroidism or hyperthyroidism can lead to significant complications, including metabolic imbalance, cardiovascular risk, and neurological symptoms. Therefore, the improved sensitivity of the proposed model is particularly valuable.

4.4 Confusion Matrix: Diagnostic Reliability

The confusion matrix indicates minimal misclassification across thyroid categories. Most predictions fall along the diagonal, suggesting accurate classification.

DOI: 10.17148/IJARCCE.2024.13254

Clinically, the few misclassifications observed are likely to occur in:

- Borderline hormone values
- Subclinical thyroid conditions

These cases are inherently challenging even for experienced clinicians, highlighting that the model's limitations align with real-world diagnostic complexity rather than systematic failure.

4.5 ROC Curve Analysis: Diagnostic Accuracy

The ROC curve demonstrates excellent performance, with an AUC close to 1.0. This indicates high sensitivity and specificity across classification thresholds.
In clinical terms:

- High sensitivity ensures detection of disease
- High specificity reduces overdiagnosis

This balance is essential in thyroid care, where both underdiagnosis and overdiagnosis carry significant consequences.

4.6 Feature Importance: Alignment with Endocrinology

Feature importance analysis identified **TSH, T3, and T4** as the most influential variables in classification.
This finding strongly aligns with established endocrinology principles:

- TSH acts as the primary regulatory hormone
- T3 and T4 reflect metabolic activity

The model's reliance on these features enhances **clinical interpretability**, making it more acceptable for real-world deployment.

4.7 Impact of Federated Learning in Clinical Context

The use of federated learning ensures that patient data remain within local institutions, addressing critical concerns related to:

- Patient privacy
- Data security
- Regulatory compliance

From a healthcare systems perspective, this enables **multi-center collaboration** without compromising confidentiality, which is particularly important in sensitive medical datasets.

4.8 Role of Lion Optimization in Clinical Accuracy

The integration of Lion Optimization improves model convergence and parameter tuning, leading to better predictive performance.
Clinically, this translates to:

- More consistent diagnostic outputs
- Reduced variability across different patient cohorts
- Improved reliability in edge cases

4.9 Overall Clinical Implications

The proposed LbFL-TDP framework demonstrates strong potential as a **clinical decision-support tool** for thyroid disorder diagnosis. Its key advantages include:

- High diagnostic accuracy and reliability
- Alignment with established endocrine knowledge
- Preservation of patient data privacy
- Robust performance across diverse populations

However, it is important to note that such models are intended to **assist, not replace, clinical judgment**. Final diagnosis should always consider:

- Patient symptoms
- Medical history
- Additional laboratory and imaging findings

4.10 Limitations and Future Scope (Clinical Perspective)

While the model performs well, certain limitations remain:

- Difficulty in distinguishing subclinical conditions
- Dependence on quality and completeness of input data
- Lack of integration with clinical symptoms and imaging

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 2, February 2024
DOI: 10.17148/IJARCCE.2024.13254

Future work should focus on:

- Incorporating longitudinal patient data
- Including additional biomarkers (e.g., antibodies)
- Integrating with electronic health record (EHR) systems

The proposed LbFL-TDP model achieved consistent and high performance across evaluation settings, with a maximum accuracy of 92% and AUC of 0.95 under 10-fold cross-validation. From a clinical perspective, this indicates reliable differentiation between euthyroid, hypothyroid, and hyperthyroid conditions. Compared to baseline models, the proposed framework demonstrated improved sensitivity and precision, reducing both missed diagnoses and false positives. Feature importance analysis revealed TSH, T3, and T4 as dominant predictors, reinforcing alignment with established endocrine diagnostic practices. The robustness of the model across folds further supports its applicability in heterogeneous clinical populations.

5. CONCLUSION

This study presents a novel federated learning framework enhanced by Lion Optimization for accurate and privacy-preserving thyroid disorder diagnosis. The proposed model achieves superior predictive performance while maintaining strong alignment with clinical endocrinology principles. By leveraging distributed datasets without compromising patient confidentiality, the framework addresses a critical challenge in healthcare AI deployment. The integration of clinically relevant biomarkers ensures interpretability, making the model suitable as a decision-support tool rather than a black-box predictor. Experimental results confirm that the proposed approach not only improves classification accuracy but also enhances robustness across diverse patient populations. Future work will focus on incorporating longitudinal patient data, expanding to additional biomarkers, and integrating the framework into real-world clinical systems. Overall, the study demonstrates that combining federated learning with metaheuristic optimization offers a promising direction for scalable, ethical, and clinically meaningful healthcare solutions.

REFERENCES

[1]. H. B. McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” *Proc. AISTATS*, 2017.
[2]. J. Konečný et al., “Federated Learning: Strategies for Improving Communication Efficiency,” *arXiv preprint arXiv:1610.05492*, 2016.
[3]. Q. Yang et al., “Federated Machine Learning: Concept and Applications,” *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, 2019.
[4]. P. Rieke et al., “The Future of Digital Health with Federated Learning,” *npj Digital Medicine*, vol. 3, 2020.
[5]. K. Bonawitz et al., “Towards Federated Learning at Scale: System Design,” *Proc. SysML*, 2019.
[6]. T. Li et al., “Federated Learning: Challenges, Methods, and Future Directions,” *IEEE Signal Processing Magazine*, vol. 37, no. 3, 2020.
[7]. A. Sheller et al., “Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data,” *Scientific Reports*, vol. 10, 2020.
[8]. S. Ramaswamy et al., “Federated Learning for Distributed Healthcare Data,” *Journal of the American Medical Informatics Association*, 2020.
[9]. A. Kumar et al., “A Comprehensive Review on Federated Learning-Based Models for Healthcare Applications,” *Artificial Intelligence in Medicine*, vol. 146, 2023.
[10]. X. Yang et al., “Review on Security of Federated Learning and Its Application in Healthcare,” *Future Generation Computer Systems*, vol. 144, 2023.
[11]. M. Joshi et al., “Federated Learning for Healthcare Domain: Pipeline, Applications and Challenges,” *ACM Transactions on Computing for Healthcare*, 2022.
[12]. Q. Li et al., “Privacy-Preserving Federated Learning: Challenges and Opportunities,” *IEEE Internet of Things Journal*, 2021.

