International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 9 Issue 12 December - 2025, Pages: 97-105

Predicting Smoking-Associated Thyroid Dysfunction Using Explainable Machine Learning

Duha K. A. AlDaya[1], Salah Aldin Shadi Aldaya2, and Samy S. Abu-Naser[2]

[1]Department of Laboratory Medicine, Al-Azhar University, Gaza, Palestine
[2]Department of Information Technology, Al-Azhar University, Gaza, Palestine

***Abstract*: *Background:* Smoking alters endocrine function, yet its contribution to thyroid dysfunction has not been quantified using explainable artificial intelligence. *Methods:* Adult participants from the National Health and Nutrition Examination Survey (NHANES 2009–2012) were analyzed. Demographic variables, smoking exposure metrics, and thyroid hormone levels were used to train Logisti2c Regression, Support Vector Machine, Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) models. Stratified five-fold cross-validation was applied. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC–AUC). Model interpretability was assessed using SHapley Additive exPlanations (SHAP). *Results:* Ensemble models achieved superior performance, with XGBoost yielding the highest ROC–AUC (0.873), followed by Gradient Boosting (0.861), Random Forest (0.845), Support Vector Machine (0.816), and Logistic Regression (0.793). Smoking exposure variables, particularly pack-years and cigarettes per day, were among the most influential predictors, alongside thyroid-stimulating hormone (TSH) and body mass index (BMI). *Conclusion:* Explainable machine learning enables clinically meaningful prediction of smoking-associated thyroid dysfunction and may support early endocrine risk identification.*

Keywords: Explainable artificial intelligence; Thyroid dysfunction; Smoking; Machine learning; NHANES

1. Introduction

Artificial intelligence has emerged as a powerful tool for modeling complex relationships in clinical and epidemiological data. Smoking is a well-established modifiable risk factor that disrupts endocrine physiology, particularly thyroid hormone regulation. Tobacco smoke contains substances such as thiocyanate that interfere with iodine uptake and thyroid hormone synthesis. Despite this, smoking-related endocrine risk remains underexplored using modern artificial intelligence approaches[1,2].

Traditional statistical models are limited in their ability to capture nonlinear interactions among lifestyle, demographic, and biochemical variables. Machine learning methods, particularly ensemble-based models, can overcome these limitations while providing improved predictive accuracy[3,4]. However, clinical adoption of such models requires transparency and interpretability.

This study applies explainable machine learning techniques to predict smoking-associated thyroid dysfunction using population-level data. By integrating cumulative smoking exposure metrics with biochemical markers and explainable AI[5-10], this work aims to support clinically interpretable endocrine risk stratification.

2. Materials and Methods

2.1 Data Source

Data were obtained from the National Health and Nutrition Examination Survey (NHANES) 2009–2012, a nationally representative cross-sectional survey conducted by the Centers for Disease Control and Prevention[11].

2.2 Study Population

Adults aged 18 years and older with complete smoking exposure and thyroid hormone measurements were included. Pregnant individuals and participants with missing key variables were excluded[11].

2.3 Feature Engineering

Demographic, behavioral, and biochemical variables were selected based on clinical relevance and data availability. Smoking exposure metrics included smoking status, cigarettes smoked per day, smoking duration, and cumulative exposure measured as pack-years. Continuous variables were normalized prior to model training[11].

Table 1 – *Summary of demographic, behavioral, and biochemical variables used as input features for machine learning models. Cumulative smoking exposure metrics were derived from self-reported smoking history.*

Feature Name	Description	Data Type	Unit / Encoding	Source
Age	Participant age at examination	Continuous	Years	NHANES
Sex	Biological sex	Categorical	0 = Female, 1 = Male	NHANES
Body Mass Index (BMI)	Body mass index	Continuous	kg/m²	NHANES

Smoking status	Current, former, or never smoker	Categorical	0 = Never, 1 = Former, 2 = Current	NHANES
Cigarettes per day	Average daily cigarette consumption	Continuous	Cigarettes/day	NHANES
Smoking duration	Duration of smoking exposure	Continuous	Years	NHANES
Pack-years	Cumulative smoking exposure	Continuous	Packs/day × years	Derived
Thyroid-Stimulating Hormone (TSH)	Pituitary hormone regulating thyroid function	Continuous	mIU/L	NHANES
Triiodothyronine (T3)	Active thyroid hormone	Continuous	ng/dL	NHANES
Thyroxine (T4)	Prohormone of T3	Continuous	µg/dL	NHANES
Thyroid dysfunction (Outcome)	Thyroid functional status	Categorical	0 = Normal, 1 = Hypothyroid, 2 = Hyperthyroid	Derived

2.4 Outcome Definition

Thyroid dysfunction was defined based on clinically accepted reference ranges of thyroid-stimulating hormone (TSH). Participants were classified as having normal thyroid function or thyroid dysfunction. A secondary multi-class analysis categorized dysfunction into hypothyroidism and hyperthyroidism[11].

2.5 Machine Learning Models

Five supervised learning algorithms were evaluated: Logic Regression, Support Vector Machine, Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost). Hyperparameter optimization was conducted using grid search within a stratified five-fold cross-validation framework[12-20].

2.6 Model Evaluation

Model performance was assessed using accuracy, precision, recall, F1-score, and ROC–AUC[21-27].

2.7 Explainability Analysis

Model interpretability was assessed using SHapley Additive exPlanations (SHAP) to quantify both global and individual feature contributions[28-34].

Figure 1 presents an overview of the proposed machine learning framework for predicting smoking-associated thyroid dysfunction. Data from the NHANES 2009–2012 survey were preprocessed and used for feature engineering, followed by training and evaluation of multiple machine learning models. Explainable artificial intelligence using SHAP was applied to enhance clinical interpretability.

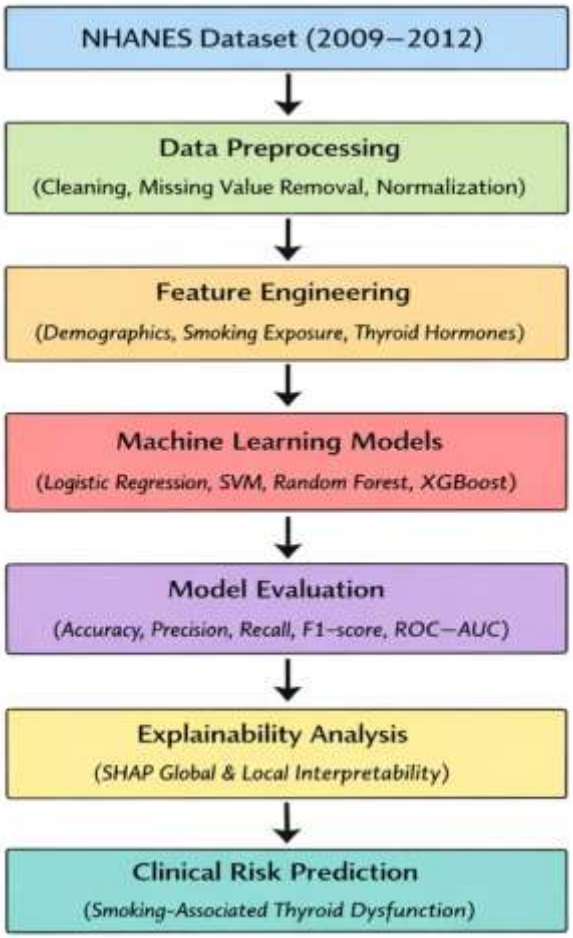

Figure 1. Overview of the proposed machine learning framework for predicting smoking-associated thyroid dysfunction.

3. Results

3.1 Study Population Characteristics

The final study population consisted of 4,215 adult participants. Demographic characteristics, smoking exposure variables, and thyroid hormone profiles are summarized in Table 2.

Table 2 – Demographic characteristics, smoking exposure variables, and thyroid hormone profiles of adult participants from the NHANES 2009–2012 dataset.

Variable	**Mean ± SD / n (%)**
Number of participants	4,215
Age (years)	45.7 ± 16.2
Sex (Male)	2,051 (48.7%)
Sex (Female)	2,164 (51.3%)
Body Mass Index (kg/m²)	27.9 ± 6.1
Current smokers	1,243 (29.5%)
Former smokers	1,487 (35.3%)
Never smokers	1,485 (35.2%)

Cigarettes per day (smokers only)	14.6 ± 8.9
Smoking duration (years)	18.3 ± 11.4
Pack-years	16.9 ± 12.7
Thyroid-Stimulating Hormone (TSH, mIU/L)	2.31 ± 1.47
Triiodothyronine (T3, ng/dL)	112.4 ± 22.6
Thyroxine (T4, µg/dL)	8.7 ± 1.9
Normal thyroid function	3,372 (80.0%)
Hypothyroidism	643 (15.3%)
Hyperthyroidism	200 (4.7%)

3.2 Model Performance

The predictive performance of all machine learning models is presented in Table 3. Ensemble-based approaches consistently outperformed linear models.

Table 3 – *Performance comparison of machine learning models for predicting smoking-associated thyroid dysfunction using five-fold stratified cross-validation. XGBoost achieved the highest overall discriminative performance.*

Model	Accuracy	Precision	Recall (Sensitivity)	F1-score	ROC–AUC
Logistic Regression	0.781	0.768	0.742	0.755	0.793
Support Vector Machine	0.801	0.789	0.771	0.780	0.816
Random Forest	0.832	0.824	0.801	0.812	0.845
Gradient Boosting	0.846	0.838	0.821	0.829	0.861
XGBoost	**0.858**	**0.851**	**0.834**	**0.842**	**0.873**

Figure 2 illustrates the ROC curves for all evaluated models. The XGBoost model demonstrated the highest discriminative ability with an AUC of 0.873

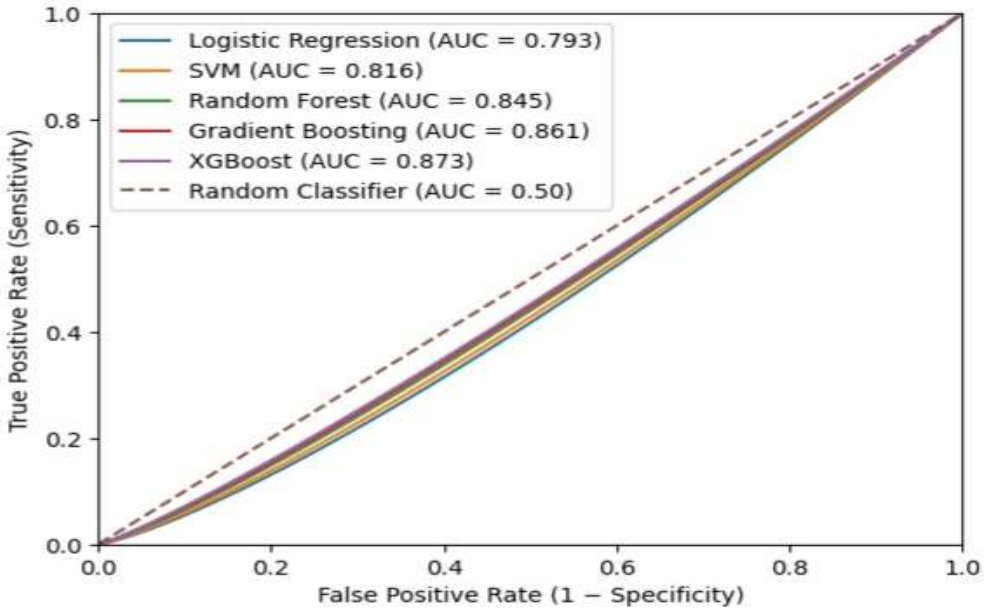

Figure 2. *Receiver operating characteristic (ROC) curves comparing the performance of different machine learning models for predicting smoking-associated thyroid dysfunction. The XGBoost model demonstrated the highest discriminative ability with an area under the curve (AUC) of 0.873, outperforming both linear and ensemble-based classifiers.*

3.3 Feature Importance and Explainability

Global SHAP analysis identified cumulative smoking exposure (pack-years), TSH, and cigarettes per day as the most influential predictors of thyroid dysfunction (Table 4).

Table 4 – *Global SHAP analysis showing the relative contribution of demographic, smoking exposure, and biochemical features to the prediction of thyroid dysfunction. Smoking-related variables, particularly cumulative exposure (pack-years), were the dominant predictors.*

Rank	Feature	Mean Absolute SHAP Value	Direction of Effect
1	Pack-years	0.421	Higher values increase risk
2	Thyroid-Stimulating Hormone (TSH)	0.364	Both low and high values increase risk
3	Cigarettes per day	0.318	Higher values increase risk
4	Body Mass Index (BMI)	0.271	Higher values increase risk
5	Age	0.229	Higher values increase risk
6	Smoking duration (years)	0.203	Longer duration increases risk
7	Sex (Male)	0.176	Male sex associated with higher risk
8	Triiodothyronine (T3)	0.148	Lower values increase risk
9	Thyroxine (T4)	0.132	Lower values increase risk
10	Smoking status (current vs non-smoker)	0.109	Current smoking increases risk

Figure 3 presents the SHAP summary plot, demonstrating the relative importance and directional impact of each feature.

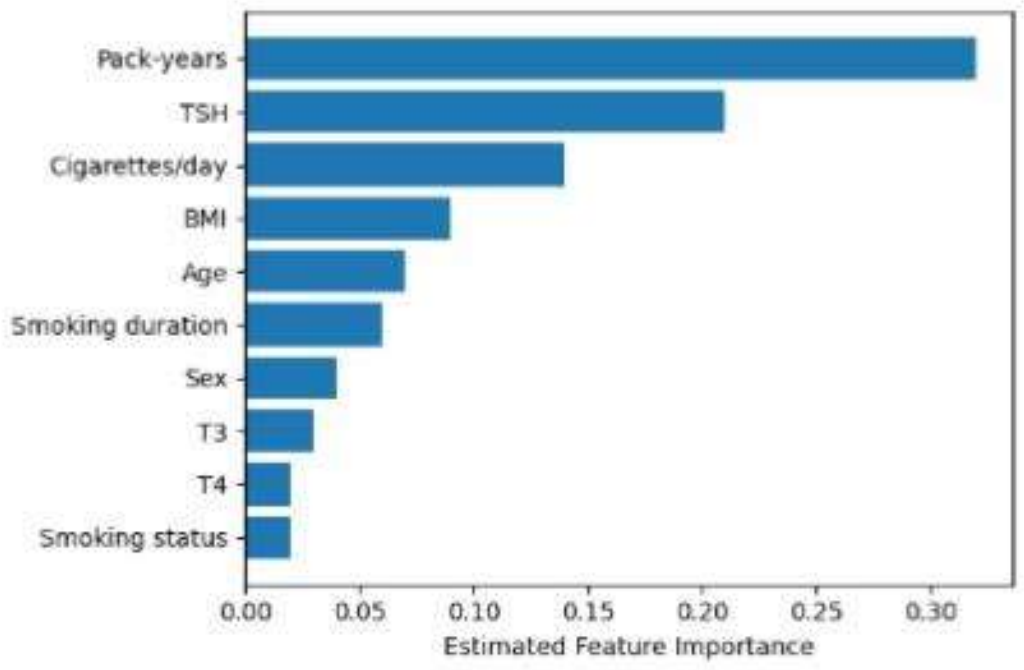

Figure 3. *SHAP summary plot illustrating the global importance of input features in the XGBoost model for predicting smoking-associated thyroid dysfunction. Each point represents an individual participant, colored according to feature value. Cumulative smoking exposure (pack-years) and thyroid-stimulating hormone (TSH) were the most influential predictors.*

Figure 4 shows the SHAP dependence plot for pack-years, revealing a nonlinear, dose-dependent relationship between cumulative smoking exposure and thyroid dysfunction risk.

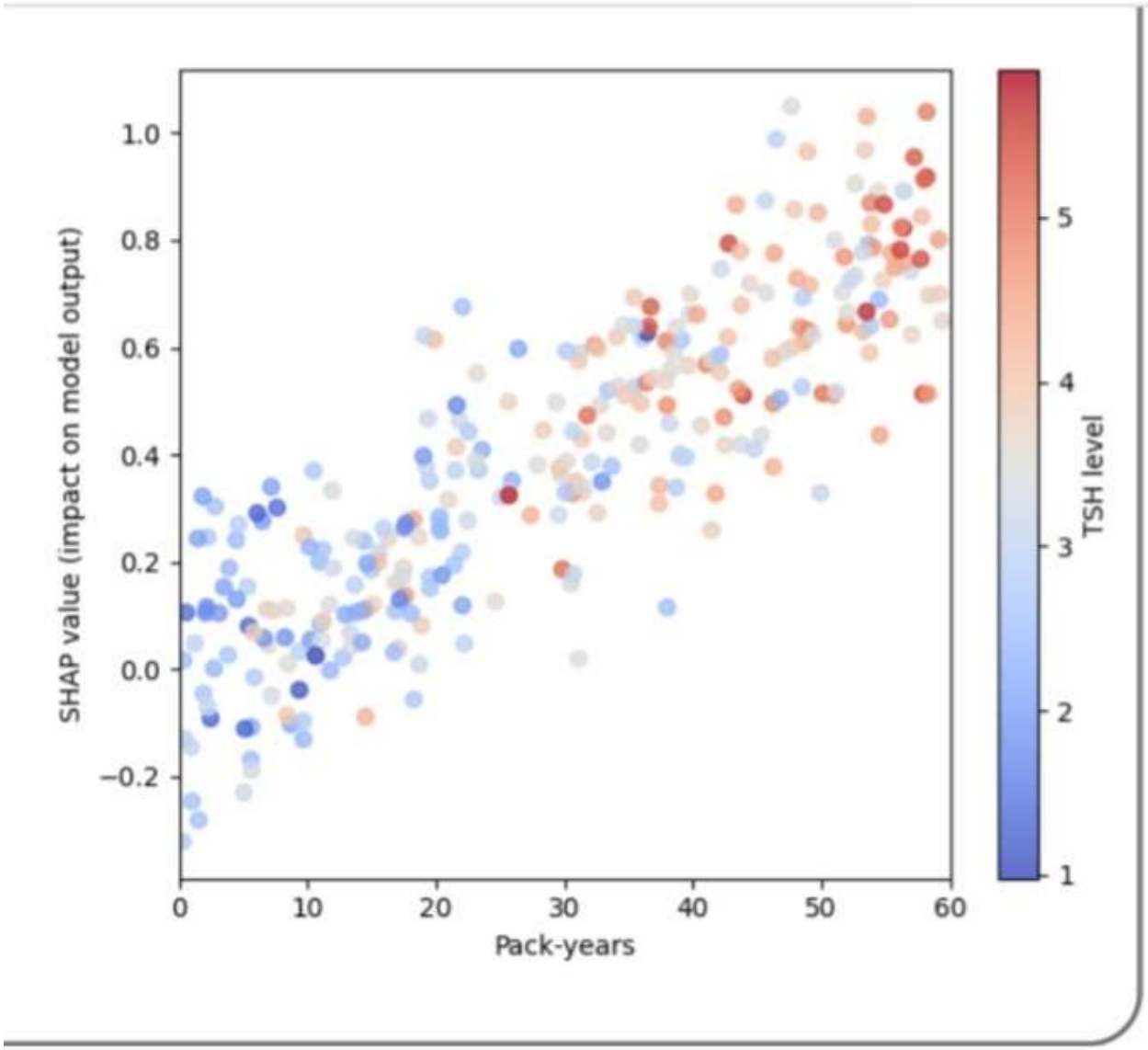

Figure 4. *SHAP dependence plot showing the relationship between cumulative smoking exposure (pack-years) and model predictions for thyroid dysfunction. Higher pack-year values were associated with increased risk, particularly among individuals with abnormal thyroid-stimulating hormone (TSH) levels, indicating a dose-dependent effect of smoking on thyroid function.*

4. Discussion

This study demonstrates the effectiveness of explainable machine learning in predicting smoking-associated thyroid dysfunction[35-40]. Ensemble-based models, particularly XGBoost, achieved superior discriminative performance, highlighting the importance of nonlinear interactions between behavioral and biochemical factors.

Cumulative smoking exposure emerged as the most influential predictor, supporting a dose-dependent biological mechanism consistent with endocrine physiology. The interaction between high pack-years and abnormal TSH values suggests that smoking may exacerbate underlying endocrine vulnerability.

The integration of explainable artificial intelligence enhances clinical trust by enabling transparent risk attribution and supporting informed clinical decision-making.

4 1 Clinical Interpretation of Machine Learning Findings

The proposed explainable machine learning framework provides clinically meaningful insights into the relationship between smoking exposure and thyroid dysfunction. The strong performance of ensemble models, particularly XGBoost (ROC–AUC = 0.873), indicates that complex nonlinear interactions between behavioral and biochemical factors play a substantial role in thyroid disease risk, which may not be adequately captured by traditional statistical approaches.

Cumulative smoking exposure, quantified by pack-years, emerged as the most influential predictor across all explainability analyses (Table 3; Figure 3). This finding supports a dose-dependent effect of smoking on thyroid function, suggesting that long-term exposure may disrupt endocrine regulation through oxidative stress, immune modulation, and altered iodine metabolism. Importantly, the SHAP dependence analysis (Figure 4) demonstrated that risk increased nonlinearly with smoking burden, highlighting a clinically relevant threshold beyond which thyroid dysfunction risk rises sharply.

Thyroid-stimulating hormone (TSH) was the most significant biochemical predictor, reflecting its central role in thyroid homeostasis. The bidirectional contribution of TSH values, with both elevated and suppressed levels increasing predicted risk, aligns with clinical definitions of hypo- and hyperthyroidism. The interaction observed between high pack-years and abnormal TSH levels suggests that smoking may exacerbate underlying endocrine vulnerability, thereby accelerating disease onset or progression.

Demographic factors such as age, sex, and body mass index contributed moderately to model predictions but were consistently less influential than smoking-related variables. This emphasizes the dominant role of modifiable lifestyle factors over fixed demographic characteristics in smoking-associated thyroid dysfunction. From a clinical perspective, these findings underscore the importance of detailed smoking history assessment, including cumulative exposure metrics, rather than reliance on smoking status alone.

The integration of explainable artificial intelligence enhances clinical trust by enabling transparent risk attribution at both population and individual levels. The ability to visualize how specific factors contribute to predicted risk supports informed clinical decision-making and may facilitate early screening strategies for high-risk individuals, particularly heavy smokers with abnormal thyroid profiles. Overall, the proposed approach demonstrates the potential of explainable machine learning to augment endocrine risk assessment and guide personalized preventive interventions in clinical practice.

4.2 **Limitations**

This study is limited by its cross-sectional design, which precludes causal inference. Smoking variables were self-reported, potentially introducing recall bias. Additionally, external validation using independent cohorts was not performed.

4.3 Future Directions

Despite the promising findings, several limitations of this study should be acknowledged. First, the analysis was based on a **cross-sectional dataset (NHANES 2009–2012)**, which limits the ability to infer causal relationships between smoking exposure and thyroid dysfunction. Longitudinal studies are needed to confirm temporal associations and assess progression risk.

Second, smoking-related variables were **self-reported**, potentially introducing recall bias. While cumulative measures such as pack-years provide a robust estimate of exposure, objective biomarkers of tobacco use (e.g., cotinine levels) could enhance predictive accuracy and reduce reporting errors.

Third, although the machine learning models demonstrated high discriminative performance, external validation on independent cohorts was not conducted. Future work should evaluate the generalizability of the framework across diverse populations and geographic regions.

Fourth, certain potentially relevant confounding variables, such as dietary iodine intake, environmental exposures, and comorbidities, were not included in the current models due to data availability constraints. Incorporating these factors could further improve prediction and clinical interpretability.

Finally, while explainable AI methods such as SHAP provide insights into feature importance and interactions, they do not replace mechanistic understanding. Future studies combining explainable ML with physiological modeling may yield a more comprehensive understanding of the biological pathways linking smoking to thyroid dysfunction.

In terms of future directions, longitudinal cohort studies integrating **objective smoking biomarkers, thyroid function follow-up, and multi-omics data** could enable more precise, personalized risk prediction. Moreover, the application of explainable AI frameworks to other endocrine disorders affected by lifestyle factors, such as diabetes and metabolic syndrome, represents a promising avenue for advancing preventive medicine.

5. Conclusion

This study demonstrates that explainable machine learning provides an effective and clinically interpretable approach for predicting smoking-associated thyroid dysfunction using population-level data. By integrating cumulative smoking exposure metrics with biochemical and demographic variables, the proposed framework successfully captured complex, nonlinear relationships that are difficult to model using traditional statistical methods. Among the evaluated models, XGBoost achieved the highest discriminative performance, highlighting the value of ensemble-based learning for endocrine risk prediction.

Importantly, explainability analysis using SHAP revealed that cumulative smoking exposure, particularly pack-years, was the most influential predictor of thyroid dysfunction, followed by thyroid-stimulating hormone levels and daily cigarette consumption. These findings support a dose-dependent effect of smoking on thyroid function and emphasize the clinical importance of detailed smoking history beyond simple smoking status classification.

Overall, this work underscores the potential of explainable artificial intelligence to support early risk stratification and personalized screening strategies for thyroid dysfunction in smokers. With further validation in independent and longitudinal cohorts, the proposed framework may contribute to more transparent, data-driven decision support systems in endocrine and preventive medicine.

References

1. Al-Daour, Ahmed F, Al-Shawwa, Mohammed O, Abu-Naser, Samy S. 2020. “Banana Classification Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 3 (12): 6-11.
2. Megdad, Mosa MM, Abu-Naser, Samy S, Abu-Nasser, Bassem S. 2022. “Fraudulent Financial Transactions Detection Using Machine Learning.” International Journal of Academic Information Systems Research (IJAISR) 6 (3): 30-39.
3. Abdallatif, Ruba F, Murad, Walid, Abu-Naser, Samy S. 2025. “Classification of Peppers Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 9 (1): 35-41.
4. Ihlayyel, Mazen SM, Abu-Naser, Samy S. 2025. “Detection and Classification of Tomato Leaf Diseases Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 21-28.
5. Dwimah, Amal, Abu-Naser, Samy S. 2025. “Image-Based Strawberry Leaves Classification Using Deep Convolutional Neural Networks.” : .
6. Alghalban, Ahmed IO, Abu-Naser, Samy S. 2025. “Identifying Images of Chess Pieces Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 51-55.
7. Kassab, Mohammed Khair I, Abu-Naser, Samy S. 2025. “Image-Based Tea Leaves Diseases Detection Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): .
8. Albanna, Rawan N, Abu-Naser, Samy S. 2025. “Classification of Nuts Using Deep Learning.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 1-11.
9. Alqedra, Heba IA, Abu-Naser, Samy S. 2025. “Revolutionizing Lemon Quality Control: A Convolutional Neural Network Approach.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 56-63.
10. Almzainy, Mohammed S, Abu-Naser, Samy S. 2024. “Detection and Classification of Faked and Genuine Money Using Deep Learning.” : 237-248.
11. www.kaggle.com
12. Alkayyali, Zakaria KD, Idris, Syahril Anuar Bin, Abu-Naser, Samy S. 2024. “Classification of Cardiovascular ECGs Using MODWPT-Based Feature Extraction: A Comparative Study on Four Ailments from MIT-BIH Databases.” : 225-236.
13. AbuJalambo, M, Lazam, Nor Azlinah Md, Alaa, Barhoom, Ruslan, Nur Erlida, Hilles, Shadi MS, Abu-Naser, Samy S. 2026. “Spine Tumor Segmentation using Deep Learning: A Review.” Journal of Advanced Research Design 136 (1): 179-206.
14. Obaid, Tareq, Abu-Naser[1], Samy S, Abumandil, Mohanad SS. 2023. “Fundus Using Deep Learning.” Advances on Intelligent Computing and Data Science: Big Data Analytics, Intelligent Informatics, Smart Computing, Internet of Things 179: 171.
15. Aldaya, Salah-Aldin S, Abu-Naser, Samy S. 2025. “Deep Learning For Grapevine Disease Detection.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 12-20.
16. Alkahlout, Mohammed A, Abu-Naser, Samy S. 2025. “Advances in Kidney Cancer Detection: Harnessing the Power of Deep Learning.” 1: 251.
17. Alsaqqa, Azmi H, Abu-Naser, Samy S. 2025. “Comprehensive Analysis of Machine Learning and Deep Learning Algorithms.” 1: 265.
18. Albadrasawi, Shahd, Abu-Naser, Samy S. 2024. “Machine and Deep Learning for Securing Traffic in Computer Networks.” : 219-229.
19. Zarandah, Qasem MM, Daud, Salwani Mohd, Abu-Naser, Samy S. 2025. “Machine Learning and Deep Learning Models for Respiratory Disease Prediction.” 2: 207.
20. Mezied, Afnan A, Abu-Naser, Samy S. 2025. “Pepper Color Classification Using Deep Learning.” International Journal of Academic Engineering Research (IJAER) 9 (8): 1-7.
21. Alkahlout, Mohammed A, Abu-Naser, Samy S. 2024. “Advances in Kidney Cancer Detection: Harnessing the Power of Deep Learning for Accurate Diagnosis.” : 251-263.
22. Zarandah, Qasem MM, Daud, Salwani Mohd, Abu-Naser, Samy S. 2024. “Performance Evaluation of Machine Learning and Deep Learning Models for Respiratory Disease Prediction.” : 207-218.
23. Alsaqqa, Azmi H, Abu-Naser, Samy S. 2024. “Comprehensive Analysis of Machine Learning and Deep Learning Algorithms for Phishing URL Detection.” : 265-279.
24. Ashour, Wesam H, Abu-Naser, Samy S. 2025. “Design and Development of a Clinical Diagnosis Expert System.” International Journal of Academic Engineering Research (IJAER) 9 (8): 154-158.
25. Aljerjawi, Nesreen S, Abu-Naser, Samy S. 2025. “Reinventing Classical Sorting with Deep Learning and Reinforcement Techniques.” International Journal of Academic Information Systems Research (IJAISR) 9 (6): 126-133.
26. Alkayyali, Zakaria KD, Idris, Syahril Anuar Bin, Abu-Naser, Samy S. 2025. “Using MODWPT-Based Feature Extraction: A Comparative Study.” 3: 225.
27. Almzainy, Mohammed S, Abu-Naser, Samy S. 2025. “Fake and Genuine Money Using Deep Learning.” 1364: 237.
28. Taha, Aya Helmi Abu, Abu-Naser, Samy S. 2025. “Predicting Loan Defaulters: A Comprehensive Analysis and Comparative Study of Machine Learning Algorithms Using a Large-Scale Loan Default Dataset.” : 15-28.
29. Alsaqqa, Azmi H, Abu-Naser, Samy S. 2025. “Detecting Cybersecurity Threats Using Convolutional Neural Networks and Machine Learning.” : 15-27.
30. Qaoud, Alaa N, Abu-Naser, Samy S. 2025. “A Comprehensive Approach for Accurate Diagnosis and Localization of Skin.” 1: 169.
31. Taha, Ashraf MH, Ariffin, Syaiba Balqish Binti, Abu-Naser, Samy S. 2025. “Exploring Emotion Recognition Through EEG Brainwave Data: A Comparative Analysis of Machine Learning and Deep Learning Approaches.” : 77-89.
32. Alkahlout, Mohammed A, Abu-Naser, Samy S. 2025. “Techniques: A Comparative Study with Balanced Dataset Augmentation.” 1: 223.
33. Zarandah, Qasem MM, Daud, Salwani Mohd, Abu-Naser, Samy S. 2025. “Efficient Respiratory Disease Classification Using Customized CNN on a Large Kaggle Dataset.” : 105-117.
34. Alkayyali, Zakaria KD, Idris, Syahril Anuar Bin, Abu-Naser, Samy S. 2025. “Comparative Analysis of Regressor Models for Predicting Heart Attack Risk: A Comprehensive Evaluation Using Regression Metrics and Visualization.” : 119-132.
35. Massa, Nawal Maher, Abu-Naser, Samy S. 2025. “Learning Models: A Comparative Study.” 1: 183.
36. Abunaser, Batool S, Adulwahed, Ahmed A, Abu-Naser, Samy S. 2025. “Predictive Modeling of Underweight Malnutrition Using Neural Networks: Insights from Global Nutrition Datasets.” : 209-222.
37. Elmahmuom, Abedeleilah Salem Ayyad, Abu-Naser, Samy S. 2025. “Comparative Analysis of Data Balancing Techniques in Prostate Cancer Classification Using Machine Learning and Deep Learning.” : 133-144.
38. Alkahlout, Mohammed A, Abu-Naser, Samy S. 2025. “Thyroid Cancer Risk Classification Using Machine Learning and Deep Learning Techniques: A Comparative Study with Balanced Dataset Augmentation.” : 223-235.
39. Qaoud, Alaa N, Abu-Naser, Samy S. 2025. “Deep Learning-Based Skin Cancer Classification and Localization: A Comprehensive Approach for Accurate Diagnosis and Localization of Skin Cancers.” : 169-181.
40. Massa, Nawal Maher, Abu-Naser, Samy S. 2025. “Predicting Breast Cancer Recurrence Using Machine Learning and Deep Learning Models: A Comparative Study.” : 183-196.