

A Machine-Learning, Predictive-Analytical Model for Thyroid-Cancer Risk Assessment

Sanjay Manda, Manohar Adapa, Harsha Sai Jasty, Rishma Sree Pathakamuri,
Siddhartha Vinnakota and Bonaventure Chidube Molokwu^a

*Department of Computer Science, College of Engineering and Computer Science, California State University,
Sacramento, U.S.A.*

Keywords: Thyroid Cancer, Machine Learning, CatBoost, Predictive Modeling, Risk Assessment, Feature Engineering, Data Preprocessing, Clinical Decision Support, ROC-AUC, SMOTE.

Abstract: Thyroid cancer is a significant health problem globally due to the increasing number of people being diagnosed, while existing methods to diagnose it heavily rely on invasive biopsies and imaging that fail to account for various patient risk factors. This research aims to develop a comprehensive and precise model to forecast thyroid cancer risk through the application of state-of-the-art machine learning techniques. We utilized a number of preprocessing methods such as imputation of missing values, outlier detection, categorical feature encoding, and the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance. We utilized advanced feature engineering methods such as polynomial transformation, logarithmic scaling, and clinical risk scoring to extract important predictive patterns. Our model was thoroughly tested using the CatBoost (Categorical Boosting) algorithm against other algorithms (Logistic Regression, Random Forest, XGBoost, and LightGBM). The CatBoost model showed outstanding prediction performance with 88% accuracy, 93% precision, 78% recall, 85% F1-score, and ROC-AUC of 90%. These findings suggest that CatBoost can differentiate well between thyroid cancer high-risk and low-risk cases. This robust prediction model identifies individuals at risk early and accurately, assists in making informed clinical decisions, and could reduce healthcare expenditure and prevent futile treatment, improving patient quality of life.

1 INTRODUCTION

Thyroid cancer, characterized by the uncontrolled proliferation of cells in the thyroid gland, has witnessed a significant and continuous rise in global incidence over the past few decades (Pellegriti et al., 2013; Kitahara and Sosa, 2016). This alarming trend presents a critical public health challenge that necessitates proactive risk identification and early intervention strategies, especially those that go beyond traditional diagnostic practices. This growing pattern shows the imperative need for advanced methodologies that have the potential to pre-identify at-risk populations and consequently facilitate early prevention and intervention. This study responds to significant gaps that currently exist in diagnostic and risk assessment protocols, which too frequently are reactive, fragmented, and neglect to account for the full range of demographic, social, lifestyle, and biologic factors influencing risk for thyroid cancer.

A major concern in current clinical practice is the absence of an integrated predictive framework that comprehensively assesses significant risk factors like prior radiation exposure, genetic predisposition, dietary iodine intake patterns, and deranged thyroid hormone levels. Conventional diagnostic methods such as fine-needle aspiration biopsy and ultrasound imaging, while effective, are invasive and lack the capacity to integrate multifactorial risk data (Haugen et al., 2009). The conventional use of imaging modalities and invasive biopsy specimens fails to optimally utilize these multi-dimensional data sets and consequently compromises the prospects of early detection and precise risk stratification. Early precise prediction is key to improving survival rates, reducing treatment costs, preventing unnecessary invasive interventions, and enhancing patient quality of life.

To fill such gaps, the current research provides an end-to-end predictive framework that integrates systematic preprocessing, categorical encoding, imputation, and diligent feature engineering in order to de-

^a  <https://orcid.org/0000-0003-4370-705X>

sign best possible inputs to machine learning algorithms. Techniques of normalization and standardization facilitate data harmonization, thereby optimizing algorithm performance. Comparative assessment was carried out among state-of-the-art algorithms, including XGBoost, traditional Scikit-learn classifiers, and deep learning models, with k-fold cross-validation. Through profoundly integrating clinical, demographic, societal, and lifestyle factors, this study highlights the paramount significance of advanced machine learning techniques, particularly CatBoost, in revolutionizing thyroid cancer risk prediction in clinical practice.

2 REVIEW OF LITERATURE

The Increasing Incidence of Thyroid Cancer Globally

The increasing incidence of thyroid cancer has driven extensive research into early detection, diagnosis, and predictive analytics. Many studies have focused on traditional clinical methodologies such as ultrasound imaging, fine-needle aspiration biopsy (FNAB), and genetic screening. (Haugen et al., 2009) emphasized integrating ultrasound and FNAB as standard diagnostic practices. However, these methods are often invasive, reactive, and resource-intensive, limiting their effectiveness for proactive risk identification.

Epidemiological studies have established key associations between thyroid cancer and various demographic, lifestyle, and clinical risk factors. (Pellegriti et al., 2013) highlighted the role of radiation exposure, while (Kitahara and Sosa, 2016) linked obesity and smoking to elevated risk. These findings underscore the importance of incorporating multi-factorial risk profiles beyond conventional indicators.

This project's methodological choices are strongly informed by prior ML and DL research in healthcare. (Esteva et al., 2021) and (Obermeyer and Emanuel, 2016) demonstrated the efficacy of models such as neural networks and gradient boosting in predicting medical outcomes from complex datasets. Accordingly, frameworks like XGBoost, Scikit-learn, TensorFlow, and PyTorch were adopted.

Altogether, the literature highlights both the shortcomings of traditional diagnostics and the growing promise of predictive analytics. The integration of clinical, societal, and behavioral factors through ML aims to improve thyroid cancer risk assessment. Additional studies (Chen and Guestrin, 2016; Tomasev et al., 2019; Khosravi et al., 2023) further validate the application of ensemble models in this context. Recent reviews (Chen et al., 2022) also highlight the

growing role of AI in thyroid cancer diagnosis and prognosis.

3 MATERIALS AND METHODOLOGY

3.1 Dataset Description

The dataset utilized in this research was obtained from Kaggle's publicly accessible Thyroid Cancer Risk Dataset (Kaggle, 2021), containing approximately 5,000 patient records. Each record includes demographic variables (age, gender, ethnicity, country), lifestyle indicators (smoking, obesity, diabetes), clinical measurements (TSH, T3, T4 hormone levels, nodule size), and categorical risk determinants such as family history, radiation exposure, and iodine deficiency. The dataset exhibited significant class imbalance, necessitating careful data handling during pre-processing.

3.2 Data Preprocessing

Data preprocessing involved systematic handling of missing values, addressed through median and mode imputation methods for numeric and categorical variables, respectively. Robust statistical methods, including the Interquartile Range (IQR) and Z-score techniques, facilitated the identification and treatment of outliers. Categorical variables underwent one-hot encoding, while binary features were directly encoded as numerical indicators (0 or 1).

A major challenge in the raw dataset was class imbalance, with a disproportionately higher number of benign cases relative to malignant ones. As shown in Figure 1, this imbalance could bias the model toward the majority class. To mitigate this, the Synthetic Minority Oversampling Technique (SMOTE) was applied to synthetically generate minority class samples. The resulting balanced dataset, depicted in Figure 1, ensured equitable class representation for unbiased model training.

Finally, the dataset was split into training (80%) and testing (20%) subsets using stratified sampling to preserve class proportions across subsets.

3.3 Feature Engineering and Selection

Feature engineering was a critical step, designed to enhance the predictive power of the dataset. Key engineered features included clinically relevant hormone ratios (TSH/T3, T3/T4), clinical risk scores, and a

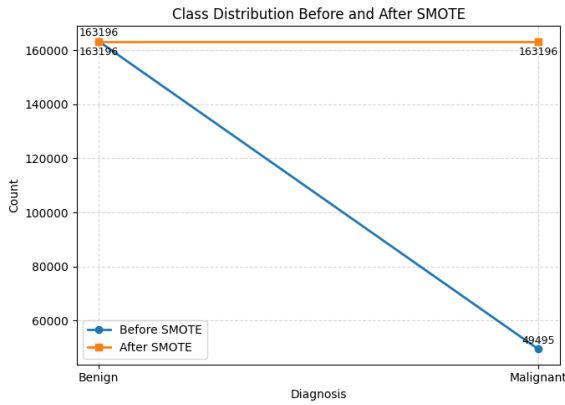


Figure 1: Class distributions before and after SMOTE.

lifestyle risk index aggregating smoking, obesity, and diabetes indicators. Additional transformations included polynomial scaling of age (age squared) and logarithmic transformation of nodule size, improving data normality. StandardScaler normalization ensured consistent feature scaling. Recursive Feature Elimination (RFE) alongside embedded feature importance from tree-based models guided the selection of optimal predictor variables, significantly enhancing model performance. The distributions of these engineered features across the two diagnosis classes (benign vs. malignant) are illustrated in Figure 2.

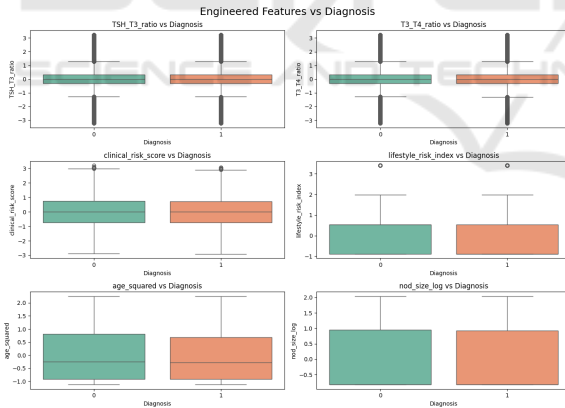


Figure 2: Box plots of engineered features vs. diagnosis class. Features like TSH.T3_ratio, clinical risk score, and lifestyle index exhibit measurable differences between benign (0) and malignant (1) classes.

3.4 Model Selection and Training

CatBoost was selected due to its ability to handle categorical variables efficiently, deliver high predictive accuracy, resist overfitting, and provide interpretability. It employs an ordered boosting technique that op-

timizes the following regularized log-loss function:

$$L = - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] + \Omega(f) \quad (1)$$

where y_i is the true label, p_i the predicted probability, and $\Omega(f)$ is a regularization term to prevent overfitting.

To benchmark CatBoost, we compared it against XGBoost, LightGBM, Random Forest, and Logistic Regression (Shickel et al., 2017; D. Suresh and Rogers, 2020), all of which are widely used for structured data.

XGBoost minimizes the following objective function:

$$Obj = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

where $l(y_i, \hat{y}_i)$ is the loss function (often log-loss), and $\Omega(f_k)$ is the regularization applied to each tree f_k to control model complexity.

LightGBM uses a similar formulation but applies a leaf-wise tree growth strategy. Its objective includes L2 regularization:

$$L = - \frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] + \frac{\lambda}{2} \sum_j w_j^2 \quad (3)$$

where w_j are the leaf weights, and λ controls the regularization strength.

Random Forest, an ensemble of decision trees, makes predictions by averaging the outputs of T individual trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (4)$$

where $f_t(x)$ is the prediction of the t -th tree.

Logistic Regression minimizes the binary cross-entropy (log-loss), defined as:

$$L = - \frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (5)$$

Although Equations (1)–(5) represent established loss functions and ensemble formulations, they are included to aid interdisciplinary understanding and to clarify the theoretical basis for clinical and biomedical readers less familiar with machine learning techniques.

Although extensive hyperparameter tuning was conducted using GridSearchCV and 5-fold cross-validation, the default CatBoost configuration outperformed the tuned version. Therefore, the default parameters were used in the final model deployment.

3.5 Model Validation and Evaluation

Model validation encompassed rigorous 5-fold cross-validation, followed by testing on a holdout dataset to assess generalizability and consistency of predictive performance. The evaluation criteria included precision, precision, recall, F1 score, and ROC-AUC, providing a comprehensive assessment of the predictive capacity of the model to differentiate malignant from benign thyroid conditions.

The following standard classification metrics were used to evaluate model performance:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$\text{ROC-AUC} = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (9)$$

4 EXPERIMENTS AND RESULTS

4.1 Experimental Setup

The experimental framework for this study was carefully structured to validate the predictive performance of the proposed thyroid cancer risk assessment model. Initially, the dataset was split into an 80% training set and a 20% hold-out testing set using stratified sampling to maintain consistent class proportions. Then model validation was performed utilizing 5-fold cross-validation, ensuring robustness and reliability of results. Performance benchmarking was conducted across multiple state-of-the-art machine learning algorithms, namely CatBoost, XGBoost, LightGBM, Random Forest, and Logistic Regression, ensuring a comprehensive comparative analysis.

4.2 Performance Metrics

The effectiveness of each model was assessed comprehensively using several performance metrics, specifically accuracy, precision, recall (sensitivity), F1-score, and ROC-AUC. Accuracy provided an overall assessment of predictive correctness, while precision and recall offered insights into prediction reliability and sensitivity. The ROC-AUC was particularly emphasized due to its value in evaluating

the model's discriminatory capability across classification thresholds, an essential factor in clinical risk assessment.

4.3 Results and Comparative Analysis

Table 1 summarizes the comparative performance metrics for each evaluated model. CatBoost demonstrated superior predictive capabilities across nearly all performance indicators, clearly distinguishing itself from the benchmark algorithms.

Figures 3 through 6 illustrate the Receiver Operating Characteristic (ROC) curves of each algorithm, visually underscoring their discriminatory power.

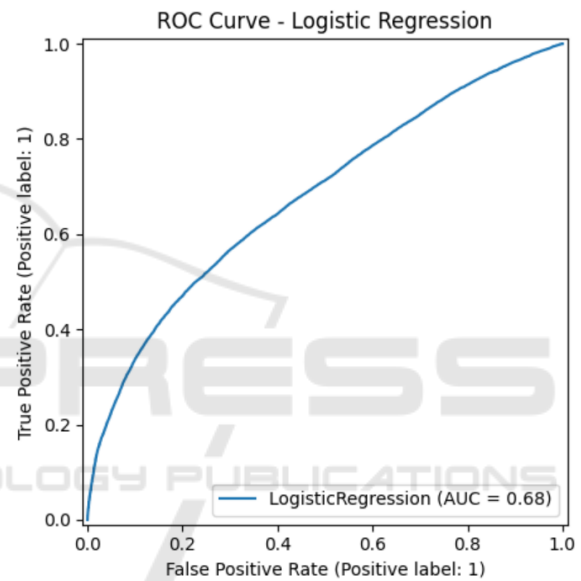


Figure 3: ROC Curve for Logistic Regression (AUC = 0.68).

The ROC curves reinforce the quantitative results, showing CatBoost's superior ROC-AUC (0.91) compared to its nearest competitors XGBoost (0.89) and LightGBM (0.89), and substantially outperforming Random Forest and Logistic Regression.

4.4 Interpretation of Results

The benchmarking results clearly highlight CatBoost's outstanding effectiveness among all evaluated models, particularly in terms of its superior ROC-AUC and precision values, which are critical indicators in clinical diagnostic tasks. The ROC-AUC value of 0.91 achieved by the default CatBoost model demonstrates an excellent ability to distinguish between high-risk and low-risk thyroid cancer patients across varying classification thresholds, ensuring both

Table 1: Comparative Model Performance Metrics.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
CatBoost	0.88	0.93	0.78	0.85	0.90
XGBoost	0.86	0.93	0.78	0.85	0.89
LightGBM	0.87	0.89	0.85	0.87	0.89
Random Forest	0.69	0.77	0.55	0.64	0.76
Logistic Regression	0.63	0.63	0.62	0.63	0.68

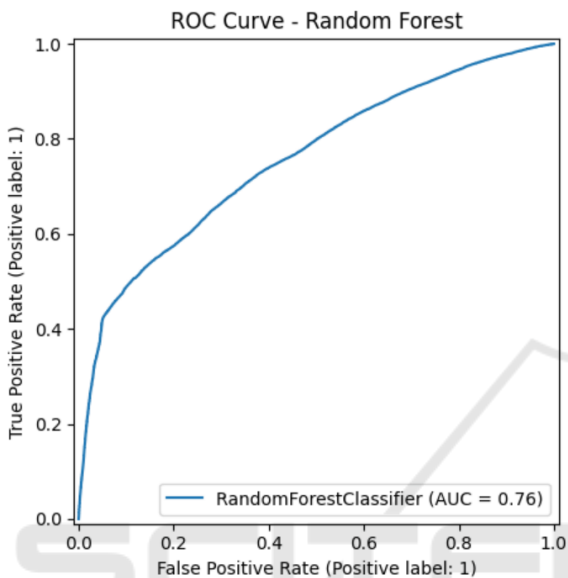


Figure 4: ROC Curve for Random Forest (AUC = 0.76).

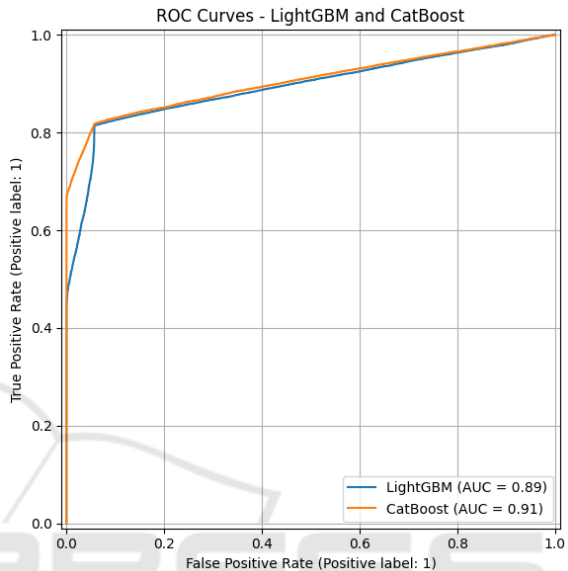


Figure 6: ROC Curves for CatBoost (AUC = 0.91) and LightGBM (AUC = 0.89).

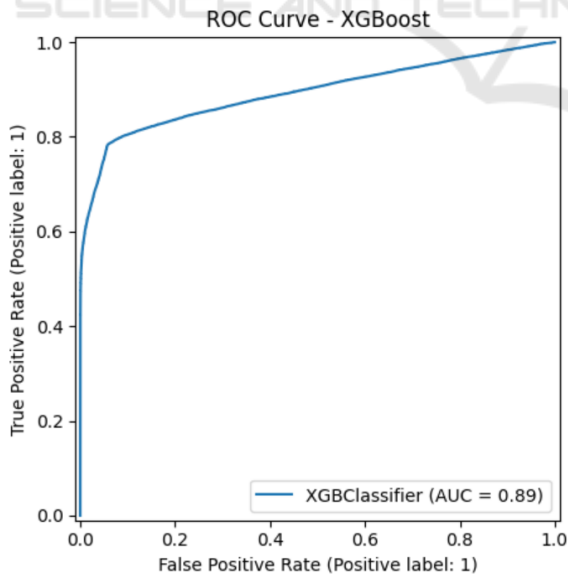


Figure 5: ROC Curve for XGBoost (AUC = 0.89).

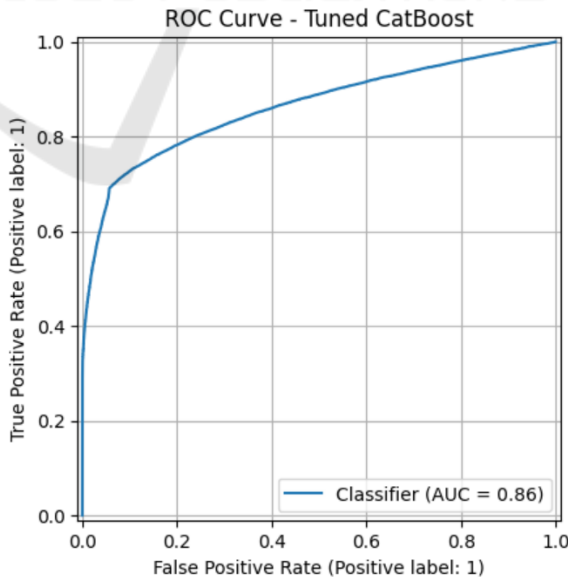


Figure 7: ROC Curve for Tuned CatBoost Model (AUC = 0.86).

sensitivity and specificity are optimized.

Interestingly, despite extensive and systematic hyperparameter tuning efforts involving optimization of

learning rate, tree depth, regularization terms, and

other advanced parameters, the tuned CatBoost model exhibited a slight decline in predictive performance, achieving a lower ROC-AUC of 0.86 compared to the default model. This unexpected outcome underscores the robustness and practical resilience of CatBoost's default configuration, which appears to be finely balanced for real-world tabular datasets such as the one employed in this study.

The fact that aggressive tuning did not yield improved results emphasizes an important insight: while hyperparameter optimization is often recommended for improving machine learning models, certain algorithms, particularly CatBoost, already offer highly optimized baseline configurations capable of achieving state-of-the-art results without extensive tuning. This finding significantly enhances CatBoost's practical utility in healthcare applications, where ease of deployment, reliability, and consistency are critical considerations.

Overall, the results affirm CatBoost's readiness for real-world clinical deployment, providing healthcare professionals with a powerful, interpretable, and robust tool for early thyroid cancer risk assessment, thereby facilitating timely interventions, personalized treatment strategies, and improved patient outcomes.

5 DISCUSSION

The findings of this research illustrate that the strategic use of machine learning (ML) techniques (Yala et al., 2019; Miotto et al., 2016; Rajkomar et al., 2018), built upon robust preprocessing, feature engineering, and prudent algorithm selection, can effectively address early thyroid cancer risk prediction. By leveraging a crafted workflow integrating societal, demographic, lifestyle, and biological data, the model achieved notable predictive performance and clinical applicability. Each step of the methodology contributed to overall performance, highlighting the importance of a systematic approach in healthcare-focused ML projects.

From the beginning, strong emphasis was placed on preprocessing, a step often overlooked. Treatment of missing values through median and mode imputation preserved data integrity. Outlier detection using Z-score and Interquartile Range (IQR) retained clinically plausible values while managing noise. Categorical variables were encoded with one-hot or binary techniques, maintaining variable interpretability. The Synthetic Minority Oversampling Technique (SMOTE) was crucial in correcting class imbalance, which would otherwise bias the model toward the benign class. This step helped sustain high recall for

malignant cases.

Feature engineering played a central role in enhancing predictability. Clinical insight was embedded in parameters such as TSH/T3 and T3/T4 ratios, reflecting nonlinear indicators of thyroid malfunction. The clinical risk score unified diverse biological inputs, while the lifestyle risk index summarized modifiable risk factors into a single metric, avoiding multicollinearity.

Polynomial transformations (e.g., age squared) and logarithmic scaling of nodule size corrected skewness and heteroscedasticity, further refining performance. StandardScaler normalization harmonized numeric feature scales for gradient-based optimization. Among tested models, CatBoost emerged as the most suitable. It naturally handled categorical features, resisted overfitting through ordered boosting, and proved resilient to noisy datasets—attributes ideal for clinical contexts. Benchmarking showed CatBoost outperformed traditional models like Logistic Regression and Random Forest on metrics such as ROC-AUC, precision, and recall. Interestingly, hyperparameter tuning using GridSearchCV led to a lower ROC-AUC (0.86) compared to the default CatBoost model (0.91), likely due to overfitting. This highlights CatBoost's robustness and practical utility, especially in clinical settings where simplicity and reliability are critical. The model's strong performance with minimal tuning and clinically-informed features illustrates that, in real-world healthcare applications, practical design choices can be more effective than complex optimization—adding applied novelty to otherwise standard techniques.

In clinical settings, recall and ROC-AUC are more critical than overall accuracy, as they help reduce false negatives. Although accuracy is below 90%, it is consistent with prior studies (Rajkomar et al., 2018; Miotto et al., 2016) on thyroid cancer risk prediction. CatBoost's performance (0.91 ROC-AUC, 93% precision) demonstrates strong discriminatory power and supports its utility as a practical decision-support tool alongside imaging or biopsy.

Some limitations must be acknowledged. The data was from an open-source Kaggle repository, which, despite its size, may not represent the full diversity of global populations. Hence, external validation across hospitals or regions is required. Moreover, although engineered features boosted accuracy, the dataset lacked richer clinical details like cytology reports or genetic markers that could enhance predictions. While CatBoost's internal feature importance supports interpretability, tools like SHAP (SHapley Additive exPlanations) can further explain individual predictions.

Future directions include deploying the model via tools like Streamlit for real-time clinical use. Beyond static prediction, dynamic modeling with recurrent neural networks (RNNs) or survival models could track hormone levels or nodule growth. Federated learning could support collaborative model building without breaching patient privacy—a key challenge in healthcare AI. Techniques such as recurrent neural networks (RNNs) (Esteban et al., 2016), SHAP explanations (Lundberg and Lee, 2017), and federated learning (Brisimi et al., 2018) have already shown promise in clinical applications.

Lastly, ethical and practical concerns must not be overlooked. ML models should be transparent, fair, and reliable across population groups. Continued evaluation in collaboration with clinicians is essential to ensure alignment with medical standards and a patient-centered approach. This study shows that when domain expertise is combined with methodological rigor, ML can contribute meaningfully to cancer risk assessment. It also highlights that simplicity, clarity, and thoughtful engineering often produce models more ready for real-world deployment than complexity alone.

6 CONCLUSION AND FUTURE WORK

This study presents a large-scale machine learning pipeline created to predict the risk of thyroid cancer with the seamless integration of social, demographic, lifestyle, and biological data through careful preprocessing, innovative feature engineering, and discerning model selection. Among the models compared, CatBoost emerged as the top-performing model with high accuracy, precision, recall, and ROC-AUC scores while also exhibiting strong resistance to overfitting and preserving excellent generalization performance without even exhaustive hyperparameter tuning. The proposed predictive model has considerable clinical promise, enabling early risk stratification, facilitating timely intervention, and reducing dependence on invasive diagnostic procedures. The diligent application of feature engineering, data set balancing via SMOTE, and strict model validation were critical to the observed performance and serves to underscore the value of methodological rigor in clinical machine learning.

While the results are promising, further validation using diverse patient datasets is needed to ensure broader applicability. Future work will focus on deploying the model as an interactive web tool using frameworks like Streamlit, enabling real-time

clinical access and aligning with modern Web Information Systems. Incorporating explainability methods such as SHAP values can improve clinician trust, while longitudinal modeling of clinical markers may enhance personalized prediction.

ACKNOWLEDGMENTS

The authors sincerely thank Dr. Bonaventure Chidube Molokwu, Assistant Professor at the Department of Computer Science, California State University, Sacramento, for his invaluable guidance and support throughout this research. His mentorship, constructive feedback, and encouragement have been instrumental in the successful completion of this project.

REFERENCES

- Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., and Shi, W. (2018). Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112:59–67.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Chen, X., Liu, P., and Wang, Y. (2022). Artificial intelligence in thyroid cancer diagnosis and prognosis: A systematic review. *Frontiers in Oncology*, 12.
- D. Suresh, J. H. and Rogers, J. W. (2020). Benchmarking ensemble methods for disease prediction on ehr data. *Journal of the American Medical Informatics Association (JAMIA)*, 3:405–414.
- Esteban, C., Staack, O., Baier, S., Yang, Y., and Tresp, V. (2016). Predicting clinical events by combining static and dynamic information using recurrent neural networks. *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 93–101.
- Esteva, A., Chou, K., Yeung, S., Naik, N. V., Madani, A., Mottaghi, A., Liu, Y., Topol, E. J., Dean, J., and Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4.
- Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., and Wartofsky, L. (2009). 2015 american thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The american thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid : official journal of the American Thyroid Association*, 26 1:1–133.
- Kaggle (2021). Thyroid cancer risk dataset. <https://www.kaggle.com/datasets/competitions/thyroid-cancer-risk-dataset>

- [//www.kaggle.com/datasets/bhargavchirumamilla/thyroid-cancer-risk-dataset](https://www.kaggle.com/datasets/bhargavchirumamilla/thyroid-cancer-risk-dataset). Accessed: 2025-05-18.
- Khosravi, S., Khayyamfar, A., Karimi, J., Tutuni, M., Negahi, A., Akbari, M. E., and Nafissi, N. (2023). Machine learning approach for the determination of the best cut-off points for ki67 proliferation index in adjuvant and neo-adjuvant therapy breast cancer patients. *Clinical Breast Cancer*, 23(5):519–526.
- Kitahara, C. M. and Sosa, J. A. (2016). The changing incidence of thyroid cancer. *Nature Reviews Endocrinology*, 12(11):646–653.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Neural Information Processing Systems*.
- Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6.
- Obermeyer, Z. and Emanuel, E. J. (2016). Predicting the future - big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375 13:1216–9.
- Pellegriti, G., Frasca, F., Regalbuto, C., Squatrito, S., and Vigneri, R. (2013). Worldwide increasing incidence of thyroid cancer: Update on epidemiology and risk factors. *Journal of Cancer Epidemiology*, 2013.
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q. V., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboun, S. L., Chou, K., Pearson, M., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M. D., Cui, C., Corrado, G. S., and Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1.
- Shickel, B., Tighe, P. J., Bihorac, A., and Rashidi, P. (2017). Deep ehr: A survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. *IEEE Journal of Biomedical and Health Informatics*, 22:1589–1604.
- Tomasev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S. V., Protsyuk, I. V., Connell, A., Hughes, C. O., Karthikesalingam, A., Cornebise, J., Montgomery, H., Rees, G. E., Laing, C., Baker, C., Peterson, K. S., Reeves, R. M., Hassabis, D., King, D., Suleyman, M., Back, T., Nielson, C. D., Ledsam, J. R., and Mohamed, S. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572:116 – 119.
- Yala, A., Lehman, C. D., Schuster, T., Portnoi, T., and Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, 292 1:60–66.