Thyroid Cancer Reoccurrence Prediction Using Machine Learning

Huthaifa Khazaleh[1] Haya Abu Al-Asal[2]
[1] Computer Information Systems, Jordan University of Science and Technology, Irbid, Jordan
[2] Computer Information Systems, Jordan University of Science and Technology, Irbid, Jordan

Abstract: Thyroid Cancer Recurrence (TCR) presents a significant concern for patients who have previously undergone treatment, as it can lead to more complicated health issues and a need for additional medical interventions. The risk of recurrence varies depending on the type and stage of the initial cancer, as well as the effectiveness of the initial treatment. Early detection of recurrent thyroid cancer is essential, as it enables prompt medical response and more effective management of the disease. Thereby improving the chances of successful treatment and minimizing the impact on the patient's quality of life. This research theorize the potential of utilizing machine learning (ML) of TCR forecasting, this research used an open source data from UCI that includes medical examinations, demographic data, medical history, and lifestyle of the patient, the research included multiple Machine learning models to get the best possible results as well as ensemble model to enhance the robustness of the model, the ensemble model used the research outperformed other models achieving 98.7% accuracy score as well as 98.7% F1-Score proving that machine learning could be useful making the TCR more efficient.

Keywords: Cancer, Machine Learning, Ensemble, Random Forest, SVM.

1 Introduction

Cancer is a major public health issues and is the second leading cause of death around the world. Thyroid Cancer (TC) is a leading cause of death, affects many body parts. Among the various types, thyroid carcinoma is one of the most common endocrine cancers worldwide, it's also considered as the most common endocrine malignancy, Concerns are growing about the rising incidence of thyroid cancer and related mortality. [1-2]

Thyroid Cancer has diverse types ranging including Frequent Papillary Tumors (FPT), aggressive follicular, medullary, and anaplastic tumors. It commonly affects women more compared to men, with incidence rate three times higher than men.[1]

Thyroid cancer is a prevalent disease but has a high (5-year) relative survival rate of 98.5%, with 2100 US recording deaths annually. Thyroid cancer diagnosis is based on a patient's medical history, including radiation doses, family history, and age. There's no known specific cause for thyroid cancer development. [3-4]

Differentiated thyroid carcinoma (DTC) accounts for 90% of cases, it's also growing rapidly threefold every decade to become the Fourth most common malignant tumor by 2030. Despite its high prevalence, it has a good prognosis and high chance of cure.[5]

Thyroid cancer diagnosis procedure included ultrasound, fine-needle aspiration cytology (FNAC), blood tests, and radioactive isotope scanning, with ultrasound being the preferred method due to its ability to avoid excessive cell puncturing. [6]

Thyroid cancer diagnosis depends on biomarker discovery to build symptomatic strategies. The best biomarker is Blood test the most convenient and difficult to contaminate blood in diagnosing diseases, and various tumor markers in blood have been widely used in diagnostic procedures, which confirm their value in diagnosing tumors. Furthermore, Overdiagnosis and overtreatment of thyroid nodules causes issues due to heterogeneity, overlapping in ultrasound images, and differences in experience. This can increase the risk of invasive testing and treatment. Early prediction of thyroid cancer disease can prevent progression, improve treatment, and improve cure rates by utilizing existing medical data. [6-7]

The main factors causing increase in TC incidence are chromosomal and genetic alterations, iodine intake, Thyroid Stimulating Hormone (TSH) level, autoimmune thyroid disease, gender, estrogen, obesity, lifestyle changes, and environmental pollutants. The gold standard for tumor detection in patients with Different Thyroid Cancer (DTC) is Tg measurement, Thyroglobulin is usually measured as T4 therapy continues (onT4-Tg), high level of thyroglobulin might be an indication of cancer recurrence. [5-8]

Machine learning is a new computer-based data analysis method, it's widely used in the medical field, especially in radiology, ophthalmology, and dermatology Compared to traditional statistical methods such as logistic regression. Machine learning enables more interactions between variables and outcomes to be found. However, to our knowledge, studies on employing machine learning for predicting Papillary Thyroid Carcinoma (PTC) are still limited. In fact, establishing a robust predictive model for PTC would help clinicians stratify high-risk patients for intensive treatment and propose candidates for active follow-up.[9]

1.2 Statement problem

Thyroid cancer is one of the most common cancers in the world and will become the fourth largest in the world in the year 2030. The fear of recurrence of thyroid cancer will remain because it is one of the diseases that can recur even after thyroid tumor has been removed and treated, and the risk of losing the patient permanently increases. Also, one of the reasons that may be an obstacle to the process of early diagnosis is that health systems around the world suffer from overcrowding and a shortage of specialists in certain specialties. This study integrated the medical sector and technology by using machine learning algorithms to improve the process of diagnosis, follow-up, and early treatment of the patient, preserve the patient's life, and improve the quality of life for patients.

1.3 Aims of the study

- Detect the potential of utilizing machine learning of TCR forecasting.
- Detecting the effectiveness of ensemble model utilization compared to traditional ML in TCR forecasting.

Research Article
Manuscript received date; accepted date; published online date

2 Related work

Reference [10] shows developed a machine learning algorithm model to predict lung metastasis of thyroid cancer using data from the National Institutes of Health's SEER database. The model included six models: SVM, logistic regression, XGBoost, decision tree, random forest, and k-nearest neighbor. The best model for predicting thyroid cancer without lung metastasis is Multivariate logistic regression. It has achieved the highest accuracy (97.9%). While it was the best model used for independent factors predicting thyroid cancer (TC) is a RF model was achieved the following accuracy (0.99%). This model can used clinical decision-making.

A study focuses on improving patient care and cancer survival prediction models. Machine learning techniques were used to create models using SEER data. Predict Thyroid Cancer prognosis of people with malignant cancer growth for predicting depend on classification techniques Seventeen critical factors were identified and classification algorithms were trained. The accuracy ranged from 97% to 99% by LR and hist gradient boosting classifier (HGB) model. The models used evaluate matrices for F1 score, precision, recall, and AUC score. The prediction model outperforms earlier research and models.[11]

A study proposed prediction for predicting the risk of thyroid disease, hyperthyroidism and hypothyroidism being common diseases. Data cleaning techniques were used to analyze and classify models for thyroid disease prediction. the study used different Machine learning techniques such as SVM, K-NN, and Decision Trees were proposed. The SVM algorithm achieved the best accuracy it was (99.63%).[12]

Reference [13] shows three novel models for predicting thyroid disease using machine learning algorithms from a dataset of 1464 Indian patients. The models using a variety of ML algorithms such as (Support Vector Machine, Naïve Byes, J48, Bagging, Boosting, and the first model achieved the highest accuracy of 98.56% with bagging. The second model achieved the highest accuracy of 99.08 with Support Vector Machine, and the third model achieved the highest accuracy of 92.07% with J48 classifier on serological tests.

A study shows for diagnosed Thyroid disease, by using machine learning algorithms. This study analyzes various classifiers, including K-nearest neighbor, Naïve Bayes, support vector machine, decision tree, and logistic regression, using data from DHQ Teaching Hospital in Pakistan. The results show that classifiers with L1-, L2based feature selection achieved higher accuracy (Naive Bayes 100%, logistic regression 100%, and KNN 97.84%) compared to those without feature selection and L2-based technique.[14]

Reference [15] shows results for prediction thyroid diseases, including hypothyroidism, hyperthyroidism, and thyroid cancer, significantly impact health. through Decision support systems (DSS) have become crucial for medical diagnosis and prediction. Traditional models struggle with interpretability, making them difficult for clinicians to trust. This study used an innovative approach to combine decision trees with ontology integration for interpretability, The model achieved accuracy (98.53%) ontology classifier and provides insights into the decision-making process. While decision tree achieved accuracy (97.85%).

A study suggested to improvement of prediction model for Thyroid disease diagnoses, this study using a deferent ML algorithm's compares Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) classifiers for diagnosing thyroid disease using a UCI dataset. The results show RF classifier performs better than SVM and K-NN, with overall classification accuracy of 98.50%, 97.02%, and 95.81% respectively.[16]

A study presented proposed system combines quantum computing with machine learning techniques to improve the accuracy for thyroid disease prediction through advanced machine learning. The study used different types of ML algorithms such as (SVM, KNN, Random Forest, QSVM). the result shows RF model achieved the worst accuracy (50.01%), but the QSVM (Quantum Support Vector Machine) more accurate classification than SVM, KNN, RF models it was achieved accuracy (98.77%).[17]

Reference [18] shows developed a machine learning prediction model for diagnosed thyroid cancer patients. Using demographic and clinicopathologic variables, this study used a random forest algorithm model that was developed. The RF model constructed in this study could accurately predict bone metastases in TC patients it was achieved (AUC: 0.917, accuracy: 0.904, recall rate: 0.833, and specificity: 0.905).

A study explores three feature selection methods for hypothyroidism, prediction, and improved performance machine learning prediction model using DT, SVM, RF, NB, and LR. are Recursive feature selection (RFE), combined with RFE and achieve 99.35% accuracy.[19]

Despite the significant advancements in Thyroid cancer and Thyroid cancer reoccurrence prediction, current research has neglected few unexplored approaches one of which is the utilization of combining the prediction strengths of multiple ML models by using ensemble techniques, to create more robust ML model. Which might make the model more versatile with reliable accuracy, precision, recall, F1-score.

This paper aims to explore the utilization of ensemble model for thyroid cancer reoccurrence, to adders the potential of ensemble technique.

3 Methodology

3.1 Methods

Support Vector Machines (SVM) are supervised learning algorithms that identify patterns. This classification approach calculates and evaluates classification functions and data.

This strategy typically outperforms other methods for categorization. A nonlinear classification approach has been reported. SVM uses a hyperplane as a decision surface to maximize the separation of positive and negative samples. This technique can take complex data and transform it into a simpler format where clear distinctions can be made. This allows for better results in tasks like classifying or predicting information. Support Vector Machines (SVMs) are a powerful method based on sound mathematical principles and have proven useful in many real-world applications, especially within the field of bioinformatics. [20]

Decision trees (DTs) are among the most powerful and commonly used classification and prediction algorithms in machine learning today. Many academics have used it as a classifier in the healthcare field to evaluate data and make decisions. DT learns decision rules from data attributes and splits data into branch-like segments to forecast a target variable's value. Input values might be either continuous or discontinuous. The leaf nodes return class labels or probability scores. It is feasible to convert the tree into a collection of decision rules. These categorization ideas can be easily demonstrated graphically. [21]

Random Forest (RF): is a data categorization strategy that employs many decision trees. Bagging and feature randomization are used to create an uncorrelated forest of trees whose committee prediction is more accurate than any individual tree commonly used in classification and regression problems. To achieve the optimum outcome, this classification technique constructs many decision trees and combines them. It mostly uses bootstrap aggregation or bagging for tree learning [21].

K-Nearest Neighbor (KNN) is a supervised classification technique that classifies new test samples based on the majority category of their K-nearest neighbors in the dataset. It categorizes data points based on their similarity to existing examples making it a versatile approach for a variety of classification problems.[22]

Logistic Regression (LR): Classification is a common application of logistic regression, a supervised learning method based on the probability function. Historically, logistic regression has been employed in the field of statistics for the purpose of analyzing data and identifying the associations between a set of independent factors and a set of dependent variables. The result of a regression analysis is a continuous numerical number. The value of a discrete dependent variable is calculated based on the value of a non-linear independent variable [23].

XGboost is a widespread end-to-end and scalable tree-boosting model. It has been employed and optimized mostly in research, and it is the enhanced structure of the gradient boosting regression model (GBRT). GBRT contains a sequence of fundamental regression trees by way of the sequential technique and accommodates multiple trees to enlarge model ability [24].

Ensemble model the pursuit of ensemble has been motivated by the intuition that an appropriate integration of different participants might leverage distinct strengths. In multiple classifier combinations, the scores generated by contributing classifiers on component feature sets are taken as inputs to the combination function [25]. In this research, multiple ensemble models have been created using multiple model combinations including using all models included ensemble model (KNN, SVM, RF, LR, DT, and XGB), and the other ensemble model utilized only (RF and XGB), but the one that the research will show and explain is the model that provides the best results among the ensemble models which showed the best results.

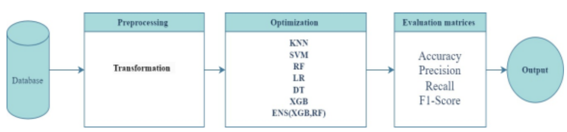

Fig.1 This figure represents the methodology of the research

The data goes throw a transformation phase before going to the model, which transform and encode the features using Label Encoder library from sklearn, then the data go throw splitting process dividing the data into training and testing datasets using the “train_test_split” library from sklearn, then the transformed data go throw the optimized models utilized in the research, the models were optimized using GridSearchCV library from sklearn to get the best possible results, the models will be evaluated based on four metrices including accuracy, precision, recall, F1-score.

3.2 Dataset

The database contains 380 medical records and included the main factors that affect thyroid cancer and is divided into seventeen categories, including Demographics (age, gender), Medical history (presence of thyroid conditions, classify risk, thyroid Function), Lifestyle habits (smoking, Hx Radiotherapy) In addition to other medical information. [26]

3.2 Performance metrics

Accuracy It is defined as the proportion of samples that are correctly classified into the overall samples and is indicated by the following equation, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative in (1), (2), and (3)

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \qquad (1)$$

Precision is defined as the computation of model identified as positive were correct to the overall count of positive prediction and is given by the following equation.

$$\text{Precision} = (TP) / (TP + FP) \qquad (2)$$

Recall is measuring the capability of a model to identify all the positive instances in a dataset. It is measured by the following equation.

$$\text{Recall} = (TP) / (TP) + (FN) \qquad (3)$$

F1-Score It is also known as the F-measure which is defined as the harmonical mean of precision and recall. It was used to identify measurements, where P is precision and R is recall.

$$\text{F1-Score} = 2 \times (P + R) / (P \times R) \qquad (4)$$

4 Result and Discussion

the models have been evaluated using precision, recall, accuracy, and F1-score, applying seven models to expand the experiment, and the results achieved by the models were as follows:

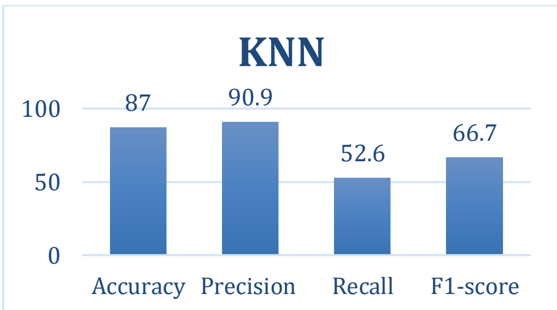

Fig.2 represents the results achieved by KNN model

The first model used in this research is the KNN model, although it achieved a decent accuracy result achieving 87% and even higher precision 90.9%, but the recall result was 52.6%, which led to low F1-score with 66.7% meaning that the KNN isn’t suitable for real world applications.

The low recall result means that the KNN model tends to miss a significant number of relevant instances (low recall). In other words, it's very cautious in making positive predictions, so it may miss out on some true positives. The cause to this situation might be the simplicity of the KNN model, which led to its struggle to capture the complexity of the underlying data, resulting in lower recall.

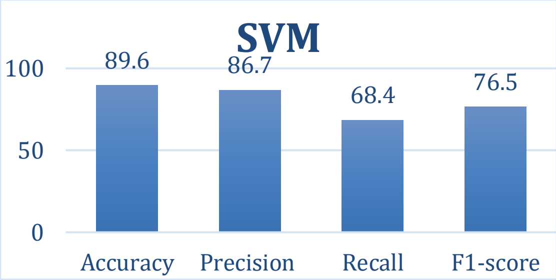

Fig.3 represents the results achieved by SVM model

The Second model used in this research is the SVM model, although it achieved a decent accuracy result achieving 89.6% and good precision 86.7%, but the recall result was 68.4%, which led to low F1-score with 76.5% meaning that even though the SVM achieved higher better results compared to the KNN, but it still isn’t suitable for real world applications.

The SVM, like the KNN, overlooks certain true positives, resulting in low recall. This may stem from the SVM's simplistic nature or from imbalanced data, which hinders its ability to grasp the intricacies of the dataset, ultimately lowering recall.

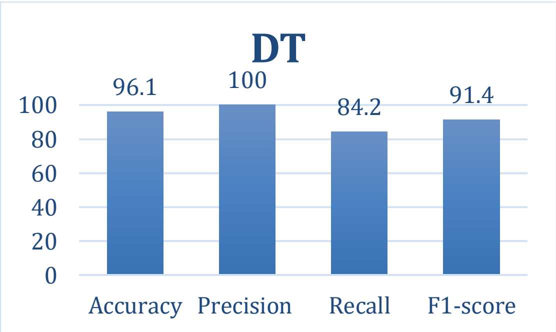

Fig.4 represents the results achieved by DT model

The third model used in this research is the DT model. It achieved good accuracy result achieving 96.1%, good precision 100%, recall result was 84.2%, which led to good F1-score with 91.4%, which is noticeable improvement over both KNN and SVM.

The decision tree algorithm showed significant enhancements across all metrics (accuracy, precision, recall, F1-Score), due to its capability to capture non-linear relationships between features and the target variable without the need for feature engineering or transformation.

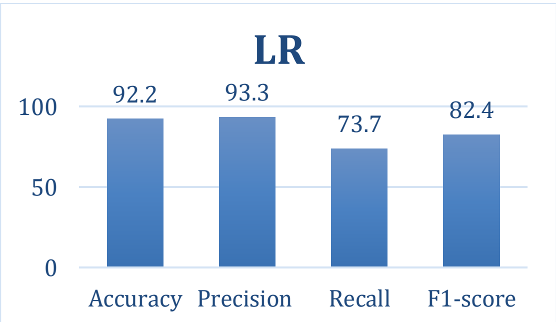

Fig.5 represents the results achieved by LR model

The fourth model used in this research is the Logistic Regression (LR) model, it achieved good accuracy result achieving 92.2%, good precision 93.3%, recall result was 73.7%, which led to F1-score with 82.4%, which is noticeable improvement over both KNN and SVM, but still lower than DT results.

While logistic regression outperformed both KNN and SVM, decision trees still exhibited superior performance across all metrics (accuracy, precision, recall, F1-Score). This could be attributed to its advantage of capturing complex relationships in the data, a limitation of logistic regression due to its assumption of a linear relationship between independent variables and the log-odds of the outcome.

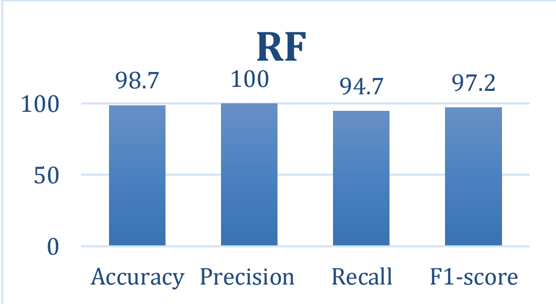

Fig.6 represents the results achieved by RF model

The fifth model used in this research is the random forest model, it achieved good accuracy result achieving 98.7%, good precision 100%, good recall result was 94.7%, which led to good F1-score with 97.2%, which is significant improvement over all models utilized in research especially the recall score.

The random forest algorithm demonstrated considerable improvements across all metrics (accuracy, precision, recall, F1-Score), attributed to its reduced susceptibility to overfitting compared to individual decision trees, due to ensemble averaging advantage.

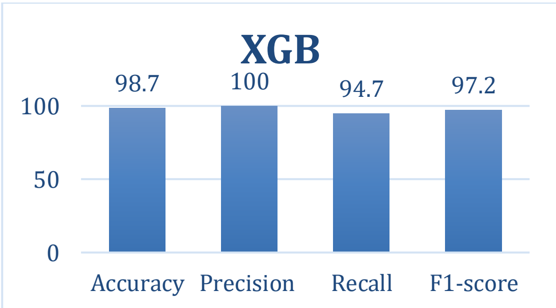

Fig.7 represents the results achieved by XGB model

The sixth model used in this research is the Extreme Gradient Boosting (XGB) model, it achieved good accuracy result achieving 98.7%, good precision 100%, good recall result was 94.7%, which led to good F1-score with 97.2%, which is similar results to the RF model.

The decision tree algorithm showed significant enhancements across all metrics (accuracy, precision, recall, F1-Score), because XGB incorporates regularization techniques to prevent overfitting, making it less prone to overfitting compared to other algorithms.

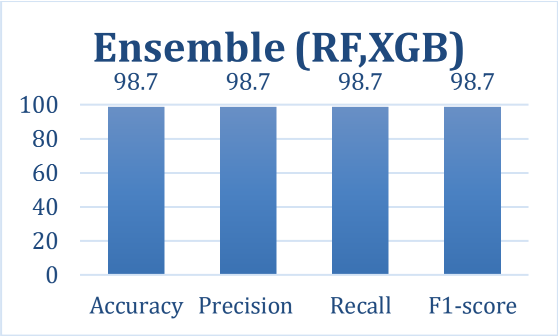

Fig.8 represents the results achieved by Ensemble model

The final model utilized in this research is the Ensemble model between XGB and RF, it achieved good accuracy result achieving 98.7%, good precision 98.7%, good recall result was 98.7%, which led to good F1-score with 98.7%, which is a good improvement in recall and F1-score results even with minor decease in precision. Overall, the Ensemble technique created the best model in terms of robustness and almost the bested all models across all categories.

5 Conclusion and future work

Many countries health systems around the world suffer from overcrowding or lack of specialists in certain sector in the health systems, technical advancement may help health systems around the world and the doctors serving in these health systems to be more efficient and precise to give patients better service and enhance their survival rate, cancer patients may suffer cancer recurrence and early detection of this recurrence will increase the chances of patient survival rate and life quality. The results of this study showed that machine learning technology has good potential to be used in medical field applications achieving high results across all evaluation metrics. Ensemble model outperformed most of the traditional models and proved to be more robust compared to RF, DT, XGB, LR, SVM and KNN.

The possibility of utilizing Deep Learning (DL) models to make better models that could use more complicated feature sets and more patient records might create more robust models that could be used in medical field applications in the future work.

Declarations of Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Author Huthaifa Khazaleh:

- Department of Computer information system, Jordan University of Science and Technology, Irbid, Jordan (Graduated):
- Email: huthaifakhazaleh@gmail.com

Haya Abu Al-Asal:

- Department of Computer information system, Jordan University of Science and Technology, Irbid, Jordan:
- Email: haabualasal224@cit.just.edu.jo

References

[1] Khazaleh, H., Al-Badarneh, A., Shannk, Y., & Al-Badarneh, A. (2021). Cancer Post-Hoc Comparative Analysis: GCC versus USA. International Journal of Biology and Biomedicine, 6.

[2] Lamartina, L., Leboulleux, S., Borget, I., & Schlumberger, M. (2022). Global thyroid estimates in 2020. the Lancet. Diabetes & Endocrinology, 10(4), 235–236. https://doi.org/10.1016/s2213-8587(22)00048-1

[3] Habchi, Y., Himeur, Y., Kheddar, H., Boukabou, A., Atalla, S., Chouchane, A., Ouamane, A., & Mansoor, W. (2023). AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions. Systems, 11(10), 519. https://doi.org/10.3390/systems11100519

[4] Hamidi, S., Hofmann, M., Iyer, P., Cabanillas, M. E., Hu, M. I., Busaidy, N. L., & Dadu, R. (2023). Review article: new treatments for advanced differentiated thyroid cancers and potential mechanisms of drug resistance. Frontiers in Endocrinology, 14. https://doi.org/10.3389/fendo.2023.1176731

[5] Olatunji, S. O., Alotaibi, S., Almutairi, E., Alrabae, Z., Almajid, Y., Altabee, R., Altassan, M., Ahmed, M. I. B., Farooqui, M., & Alhiyafi, J. (2021). Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset. Computers in Biology and Medicine, 131, 104267. https://doi.org/10.1016/j.compbiomed.2021.104267.

[6] Garo, M. L., Campennì, A., Petranovic-Ovcaricek, P., D'Aurizio, F., & Giovanella, L. (2023). Evolution of thyroid cancer biomarkers: from laboratory test to patients' clinical management. Clinical Chemistry and Laboratory Medicine (CCLM), 61(5), 935-945.

[7] Jia, Z., Huang, Y., Lin, Y., Fu, M., & Sun, C. (2023). Multidimensional prediction method for thyroid cancer based on spatiotemporally imbalanced distribution data. IEEE Access.

[8] Wang, W., Chang, J., Jia, B., & Liu, J. (2020). The blood biomarkers of thyroid cancer. Cancer management and research, 5431-5438.

[9] Crnčić, T. B., Tomaš, M. I., Girotto, N., & Ivanković, S. G. (2020). Risk factors for thyroid cancer: What do we know so far? Acta Clinica Croatica. https://doi.org/10.20471/acc.2020.59.s1.08.

[10] Lai, S., Fan, Y., Zhu, Y., Zhang, F., Zheng, G., Wang, B., Wan, Z., Liu, P., Yu, N., & Qin, H. (2022). Machine learning-based dynamic prediction of lateral lymph node metastasis in patients with papillary thyroid cancer. Frontiers in Endocrinology, 13. https://doi.org/10.3389/fendo.2022.1019037

[11] Alhashmi, S. M., Polash, M. S. I., Haque, A., Rabbe, F., Hossen, S., Faruqui, N., ... & Abubacker, N. F. (2024). Survival Analysis of Thyroid Cancer Patients Using Machine Learning Algorithms. IEEE Access.

[12] Tyagi, A., Mehra, R., & Saxena, A. (2018, December). Interactive thyroid disease prediction system using machine learning technique. In 2018 Fifth international conference on parallel, distributed and grid computing (PDGC) (pp. 689-693). IEEE.

[13] Mir, Y. I., & Mittal, S. (2020). Thyroid disease prediction using hybrid machine learning techniques: An effective framework. International Journal of Scientific & Technology Research, 9(2), 2868-2874.

[14] Abbad Ur Rehman, H., Lin, C. Y., Mushtaq, Z., & Su, S. F. (2021). Performance analysis of machine learning algorithms for thyroid disease. Arabian Journal for Science and Engineering, 1-13.

[15] OUARTANI, S., & TALEB, N. (2024).Decision Support System for Thyroid Disease Prediction Using Decision Tree Algorithm And Ontology.

[16] Shahid, A. H., Singh, M. P., Raj, R. K., Suman, R., Jawaid, D., & Alam, M. (2019, July). A study on label TSH, T3, T4U, TT4, FTI in hyperthyroidism and hypothyroidism using machine learning techniques. In 2019 International conference on communication and electronics systems (ICCES) (pp. 930-933). IEEE.

[17] Sha, M. (2023). Quantum intelligence in medicine: Empowering thyroid disease prediction through advanced machine learning. IET Quantum Communication.

[18] Liu, W. C., Li, M. X., Wu, S. N., Tong, W. L., Li, A. A., Sun, B. L., ... & Liu, J. M. (2022). Using machine learning methods to predict bone metastases in breast infiltrating ductal carcinoma patients. Frontiers in public health, 10, 922510.

[19] Gupta, P., Rustam, F., Kanwal, K., Aljedaani, W., Alfarhood, S., Safran, M., & Ashraf, I. (2024). Detecting thyroid disease using optimized machine learning model based on differential evolution. International Journal of Computational Intelligence Systems, 17(1),

3.
[20] Yilmaz, R., & YAĞIN, F. H. (2022). Early detection of coronary heart disease based on machine learning methods. Medical Records, 4(1), 1-6.
[21] Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. Iraqi Journal For Computer Science and Mathematics, 4(2), 31-42.
[22] Eldora, K., Fernando, E., & Winanti, W. (2024). Comparative Analysis of KNN and Decision Tree Classification Algorithms for Early Stroke Prediction: A Machine Learning Approach. Journal of Information Systems and Informatics, 6(1), 313-338
[23] Arif, Z. H., & Cengiz, K. (2023). Severity classification for COVID-19 infections based on lasso-logistic regression model. International Journal of Mathematics, Statistics, and Computer Science, 1, 25-32.
[24] Alghazzawi, D. M., Alquraishee, A. G. A., Badri, S. K., & Hasan, S. H. (2023). ERF-XGB: Ensemble random forest-based XG boost for accurate prediction and classification of e-commerce product review. Sustainability, 15(9), 7076.
[25] Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information sciences, 181(6), 1138-1152.
[26] UCI Machine Learning Repository. (n.d.)
https://archive.ics.uci.edu/dataset/915/differentiated+thyroid+cancer+recurrence.
[27] Barfungpa, S. P., Samantaray, L., Sarma, H. K. D., Panda, R., & Abraham, A. (2023). Dt-SNE: Predicting heart disease based on hyper parameter tuned MLP. Biomedical Signal Processing and Control, 86, 105129.
[28] Soni, V. D. (2020). Chronic disease detection model using machine learning techniques.International Journal of Scientific & Technology Research, 9(9), 262-266 .