Enhancing Thyroid Cancer Diagnostics Through Hybrid Machine Learning and Metabolomics Approaches

Meghana G Raj
Department of School of Computer Engineering,
Kalinga Institute of Industrial Technology (Deemed to be University), Bhubaneswar, India

***Abstract*—Thyroid cancer, a prevalent endocrine malignancy, necessitates advanced diagnostic techniques for accurate and early detection. This study introduces an innovative approach that integrates hybrid Machine Learning (ML) algorithms with metabolomics, offering a novel pathway in thyroid cancer diagnostics. Our methodology employs a range of hybrid ML models, combining the strengths of various algorithms to analyze complex metabolomic data effectively. These models include ensemble methods, neural network-based hybrids, and integrations of unsupervised and supervised learning techniques, tailored to decipher the intricate patterns within metabolic profiles associated with thyroid cancer. The study demonstrates how these hybrid ML algorithms can efficiently process and interpret metabolomic data, leading to enhanced diagnostic accuracy. By leveraging the distinct characteristics of each ML model, our approach not only improves the detection of thyroid cancer but also contributes to a deeper understanding of its metabolic underpinnings. The findings of this study pave the way for more personalized and precise medical interventions in thyroid cancer management, showcasing the potential of hybrid ML models in revolutionizing cancer diagnostics. Our system analyzes thyroid cancer metabolomic data using ensemble methods, neural network-based hybrids, and unsupervised and supervised learning integrations. The research shows hybrid ML models may revolutionize cancer diagnoses by improving accuracy. LSTM+CNN, LSTM+GRU, and CNN+GRU have high accuracy rates, helping us comprehend thyroid cancer's biochemical roots. Hybrid ML models enhance thyroid cancer diagnosis and management, enabling more tailored and accurate medical treatments. The hybrid machine learning models like LSTM+CNN, LSTM+GRU, and CNN+GRU beat CNN, VGG-19, Inception-ResNet-v2, decision support, and random forests (99.45%).**

Keywords*—Thyroid cancer; hybrid ML models; metabolomics; diagnostic accuracy

I. Introduction

AI improves diagnosis, treatment, and care. AI's pattern recognition, predictive analysis, and decision-making skills enable computers to analyze complicated medical data with unprecedented precision and scale [1, 2]. This discovery enhances early sickness detection, precise diagnosis, and individualized therapy. AI technologies enhance hospital operations, predict disease outbreaks, and significantly improve patient outcomes. AI is critical to provide equitable access to high-quality treatment across geographic boundaries. As AI advances, it will improve global health outcomes with increasingly complex healthcare applications. However, healthcare AI adoption is hard. User adoption of AI-driven help requires trust. Studying security, risk, and trust on healthcare, AI adoption shows that trust is crucial. Oncology's leading killer affects several organs [3, 4]. Thyroid carcinoma is a prevalent endocrine malignancy worldwide. The sixth most prevalent cancer in women aged 15–49 is thyroid cancer, hence better identification and treatment are needed (see Fig. 1). Thyroid cancer is becoming more common, and machine learning and metabolomics may enhance detection and therapy [5, 6, and 7].

Thyroid cancer has increased in recent decades, with the American Cancer Society expecting 43,800 new cases and 2,230 fatalities in 2022 [8]. Thyroid cancer develops as a nodule at the throat's base when cells proliferate rapidly and escape the immune system. Unregulated cell reproduction spreads rogue cells into surrounding tissues. About 95% of thyroid malignancies are follicular or papillary. Effective management and damage reduction require early discovery and treatment of malignant thyroid nodules. Early thyroid cancer screening detects cancerous nodules. Neck palpation during physical examinations and ultrasonography, which may detect nodules smaller than 1 cm, are the main detection modalities. Ultrasonography helps distinguish benign from malignant nodules by their features [9].

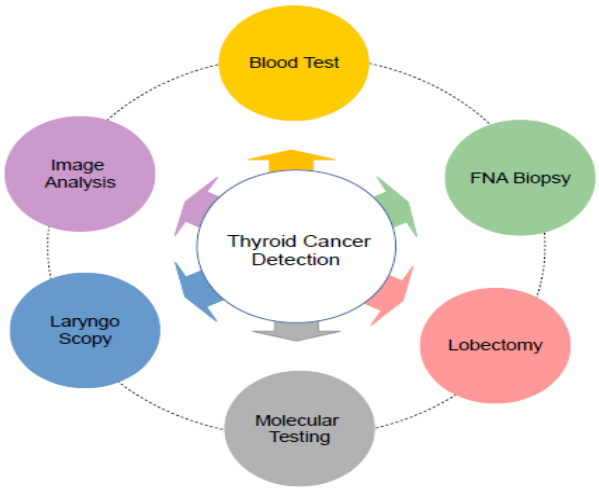

Fig. 1. Various methods for detection of thyroid cancer.

Automated thyroid nodule identification using computer-aided diagnostic (CAD) has evolved in recent years [10]. CAD tools using artificial intelligence analyze ultrasound features more intelligently, accurately, and consistently. This helps decrease needless biopsies. Machine learning and deep learning, key components of AI-based CAD systems, have changed medicine. These approaches use expert knowledge to choose important attributes from predetermined region-of-interest criteria. Margin, form, echogenicity, calcifications, and composition in thyroid ultrasound images have helped build CAD systems. Support vector machines, GoogLeNet, and CNNs have transformed thyroid nodule detection, according to previous studies. Machine learning and AI have greatly improved the use of CAD tools in clinical practice [11].

In thyroid cancer, early and accurate detection may save lives. This cancer survives better with early detection and treatment [12]. Early detection may reduce benign tumor treatment costs and stress. Doctors can enhance patient care with fast and appropriate treatment. Cancer detection employs image processing, deep learning, and AI, notably in medical imaging. Reduce noise in ultrasound images to identify thyroid cancer. Next, segmentation separates cancer-prone regions. Cancerous or benign nodules are determined by these sites. First, gather ultrasound images, then segment them to focus on the affected area. These segments' attributes constitute a predictive model, and a classifier predicts [13]. The neck gland's thyroid carcinoma is treated better with early detection. Healthcare professionals use machine learning algorithms to handle pandemics and natural disasters [14]. These algorithms help physicians identify and treat patients by analyzing enormous medical data. Thyroid CAD systems must be precise to minimize delays or unnecessary treatments. Deep learning-enhanced ultrasonography detects thyroid cancer using complex acoustic features. CAD and AI enhance thyroid cancer diagnosis. They simplify ultrasound-based risk categorization and enhance thyroid nodule identification and assessment. Traditional diagnostic methods like FNAB are 20% incorrect. Machine learning improves insights and judgments using probability and statistics. ML-based classification models using large image datasets are promising for study. Progress is shown by statistical pattern identification and quantification algorithms in thyroid node categorization systems like AmCAD-UT [15].

In this paper, the advent of deep learning models, particularly Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), and their hybrid combinations like LSTM+CNN, LSTM+GRU, and GRU+CNN, presents a transformative approach in thyroid cancer diagnostics. These advanced computational models offer unparalleled precision in analyzing complex medical data, significantly enhancing our ability to detect and diagnose thyroid cancer early and accurately. As we continue to integrate these sophisticated AI methodologies into medical practices, the potential for improving patient outcomes and revolutionizing the field of oncology is immense.

II. Related and Recent Work

AI has substantially improved medical diagnostics, notably tricky thyroid disorders. Improved ultrasound picture interpretation and quicker processing are the main reasons. Ultrasonography, FNA, and thyroid surgery now utilize deep learning (DL) and machine learning (ML) to classify thyroid nodules automatically [16]. Many research have shown AI's potential in cancer detection, where data volume and classification accuracy are critical. Ultrasound, CT, MRI, radioactive iodine, and histopathology diagnosis thyroid cancer. Many research have built AI-based CAD models to detect thyroid abnormalities in ultrasound and histopathological pictures. Xu et al. created a contrast-enhanced thyroid ultrasound diagnostic model using CNN feature extraction and LSTM classification. Zhao et al. offered CNN-extracted characteristics and image texture for ultrasonography thyroid classification [17]. CNNs classify thyroid and breast cancer ultrasound images, whereas U-Net models segment thyroid ultrasounds. Additionally, multi-scale region-based detection networks like Resnet50 and ZFnet are more accurate. Transfer learning reduces overfitting in thyroid nodule classification models employing inception networks, VGG16, and GoogLeNet. Using simple CNN models and spatial and frequency domains, Nguyen et al. categorized the TDID dataset using voting ensemble [18]. In ensemble learning, hunger games search algorithm and D-CRITIC TOPSIS model ranking educated deep vision Transformer and Mixer models. Sun et al. employed the TC-ViT model, a vision transformer with contrast learning, to classify thyroid imaging data by TI-RADS scores [19].

New AI technologies may standardize and enhance uncertain thyroid nodule categorization. Digital thyroid fine needle aspiration biopsy images are employed in these studies. EfficientNetV2-L image classification works in thyroid fine needle aspiration cytology, according to Hirokawa et al [20]. Kezalarian [21] studied AI's role in follicular cancer vs. adenoma, whereas Alabrak et al. proposed a CNN model with good accuracy, sensitivity, specificity, and AUC-score [22].

AI-based thyroid pathology whole slide image analysis utilizing modified QUADAS-2 was evaluated by Girolami et al [23]. Using numerous histopathology pictures, Wang et al. trained VGG-19 and Inception-ResNet-v2 models to diagnose thyroid diseases [24]. Chandio et al. suggested a CNN-based MTC detection decision support system. Hossiny et al. correctly identified thyroid tumors using cascaded CNN and split classification [25]. Do et al.'s thyroid cancer MI Inception-v3 model improved classification accuracy [26]. Bohland et al. found feature-based and deep learning-based thyroid carcinoma classification equivalent [27]. Transformer and Mixer models improve vision, but thyroid feature extraction is uncertain. Vector redundancy may cause feature extraction overfitting. Espadoto et al.'s dimensionality reduction survey was impressive [28]. Meta-heuristic feature selection approaches like moth flame and cuckoo optimization are common for high-dimensional datasets. In data-limited medical disciplines, ensemble techniques using numerous weak learners increase classification model accuracy. The weighted average ensemble technique is intriguing, but weight selection is tricky. FOX optimization, which performs well in traditional benchmarks, holds potential in feature selection and ensemble learning but has not been implemented [14].

TABLE I. COMPARATIVE STUDY ON VARIOUS METHODS FOR DETECTION OF THYROID CANCER

Authors	Method	Accuracy	Key Contributions
Alabrak et al. 2023 [22]	CNN model	78%	Proposed a CNN model to classify thyroid cancer with good accuracy, sensitivity, specificity, and AUC-score.
Wang et al. 2019 [24]	VGG-19 and Inception-ResNet-v2 models	97.34% and 94.42%	Trained models to diagnose thyroid diseases using histopathology images.
Chandio et al. 2020 [39]	CNN-based decision support system	99.00%	Suggested a system for detecting medullary thyroid cancer using CNN.
Hossiny et al. 2021 [25]	Cascaded CNN and split classification techniques	98.74%	Identified thyroid tumors with high accuracy.
Bohland et al. 2021 [27]	Feature-based and deep learning-based models	89.70% (feature-based), 89.10% (deep learning-based)	Comparison of thyroid tumor classification models.
Kouznetsova et al. (2021) [34]	ML model using saliva metabolites	Not Specified	Differentiated between malignant oral lesions and periodontitis.
Cai et al., 2015 [35]	Random forest ML model	86.54%	Classified lung cancer using DNA methylation markers.

The American Thyroid Association endorsed intraoperative frozen sections (FSs) for classical papillary thyroid cancer detection in 2015. However, onsite pathologists may struggle to detect rare cancers and poorly prepared specimens using paraffin sections. PTC is one of the most frequent thyroid cancer, although follicular, medullary, and undifferentiated carcinomas stain poorly. Rare lung and breast cancers are hard to diagnose. Diagnostic discrepancies and CNN model building are difficult due to the lack of pathological imaging data from uncommon cancers [29]. Computational pathology must find rare or intermediate groupings. Deep learning in several fields, including CNNs and RNNs, has led to computer-aided histopathological diagnostic systems. Digital pathology allows histopathological diagnosis using deep learning algorithms. Thanks to CAMELYON16 and the TCGA, patch-based CNNs for whole-slide images (WSIs) have improved cancer histology [30]. CNN methods for breast cancer lymph node metastatic diagnosis are examined. InceptionV3, utilizing CAMELYON16, achieves 98.6% AUC and 87.3% FROC. WSI patch image analysis using Resnet and conditional random fields was also helpful [31]. CNNs are cancer-trained and tested. On TCGA non-small cell lung cancer histopathology pictures, InceptionV3 and pathologists fared similarly. Deep learning model interpretability has improved with new bladder cancer and other cancer screening methods. Positive pathologist diagnostic accuracy comparisons [32].

ML has been used for about two decades to diagnose and track cancer. Cancer diagnosis has relied on decision trees and ANNs since the mid-1980s [33]. Age, health, sickness kind, location, tumor grade, and size affect cancer prognosis. ML predicts nodes and patient severity using this data. Protein markers and microarray data are used in breast and prostate cancer research to identify cancer types, predict risk, and test patients. Diagnostics improve using ML models for CT scans and cancer image projection. ML models to aid doctors in these imaging methods were recognized by the NCI (2022). Metabolomics has been used in cancer research. Research into cancer metabolites has advanced. In bladder cancer (BCa) studies, ML compared metabolite patterns at different stages. Metabolites in healthy and oral cancer patients are linked via these pathways. Using ML, Kouznetsova et al. (2021) distinguished malignant from periodontitis oral lesions using saliva metabolites [34]. Genomic data was used by ML-based classifiers to construct a lung cancer DNA methylation indicator panel. A random forest ML model diagnosed lung cancer with 86.54% accuracy by Cai et al., 2015 [35]. These results show ML's growing role in cancer diagnostic accuracy and efficiency, paving the way for future research (see Table I).

III. METHODOLOGY

A. Long Short-Term Memory (LSTM)

The Recurrent Neural Network (RNN) LSTM may learn long-term data sequence relationships. This is valuable in medical diagnostics, as patient data covers extended periods and includes important sequential patterns. The input, forget, and output gates make up LSTM. These gates regulate cell state information flow, enabling the network to remember or forget [36, 37]. This approach uses LSTMs to evaluate and understand complicated metabolic data and other temporal information. Metabolic marker alterations, thyroid symptom progression, and diagnostic data integration are included. The input, forget, and output gates of LSTM control information flow, making it effective. These gates determine what data to keep or discard as the data sequence proceeds, allowing the model to keep important data and forget non-essential data [40].

Equations for LSTM:

Forget Gate selects cell state data to discard. It examines the previous state (h_{t-1}). The program produces a value between 0 and 1 for each cell state (C_{t-1}) and current input (x_t).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + d_f)$$

Input Gate: A sigmoid layer chooses values to update while a tanh layer creates a vector of candidates.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)$$
$$C'_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$$

The previous cell state is updated to the new one.

$$C_t = f_t * C_{t-1} + i_t * C'_t$$

Output Gate: It selects the next hidden state representing prior inputs.

$$O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o)$$
$$h_t = O_t * \tanh(C_t)$$

LSTM can analyze consecutive metabolic profiles and historical patient data to capture temporal relationships needed for thyroid cancer diagnosis and prognosis. Combining LSTM with additional machine learning methods like CNNs for image analysis improves thyroid cancer diagnosis.

B. Gated Recurrent Units (GRU)

GRUs are Recurrent Neural Networks (RNNs) intended to analyze data sequences. It is fewer gates and are simpler and more efficient in certain cases. Medical diagnostics use GRUs to capture temporal connections in sequential data. GRUs might assess time-dependent changes in metabolites, hormone levels, and other biochemical indicators of thyroid disorders to diagnose thyroid cancer [38].

Integrating the "forget and input" gates into one GRU simplifies RNNs "update gate." They also use a "reset gate." These two gates in GRUs control the information flow within the unit, which is essential for maintaining relevant information over different time steps.

Equations for GRU:

Update Gate: Determines how much previous knowledge to pass on.

$$z_t = \sigma(W_z[h_{t-1}, x_t])$$

Reset Gate: Decides how much of the previous knowledge to forget.

$$r_t = \sigma(W_r[h_{t-1}, x_t])$$

Current Memory Content: **Creates** the candidate which will be used to update the cell state.

$$h'_t = \tanh(W.[r_t.h_{t-1}, x_t])$$

Final Memory at Current Time Step: **Combines** the old state with the new candidate state

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t$$

GRUs' basic design lets them store key data from prior data points and reject irrelevant data, improving their sequential data prediction abilities. Instead of update and reset gates, GRUs use a simpler gating method. The gates regulate how much previous knowledge to pass on to the future, improving sequential data learning. For a complete examination, we will use GRUs and other machine learning methods like CNNs.

C. Convolutional Neural Networks (CNN)

CNNs are powerful deep neural networks for visual analysis. They excel in automatically detecting and learning spatial hierarchies of characteristics from pictures, which is essential for thyroid cancer detection in medical imaging [39]. Convolutional, pooling, and fully linked layers make up a CNN. The output of a convolution process is sent to the next layer. The network builds a learnt feature hierarchy this way.

Equations for CNN:

Convolution Operation:

$$F_{ij} = \sum_m \sum_n I_{(i+m)(j+n)} K_{mn}$$

where, F_{ij} is the output feature map, I represents the input image, and K is the kernel or filter applied to the image.

Activation Function (ReLU): $f(x) = \max(0, x)$ used to provide the model non-linearity to learn more complicated patterns.

Pooling Operation (Max Pooling): $P_{ij} = max(I_{(i:i+p)(j:j+p)})$ where P_{ij} is the output after pooling, and I the input feature map, and p the pooling window size. As in ordinary neural networks, neurons in the final layers are completely coupled to all activations in the preceding layer. This part is typically used to classify features learned by the CNN into different categories.

This approach analyzes thyroid ultrasound pictures using CNNs. Their capacity to extract and learn key elements from these photos is critical for thyroid cancer detection. CNN and LSTM or GRU will evaluate non-imaging data together. CNNs extract features from pictures, whereas LSTM/GRU models analyze consecutive patient histories and metabolic profiles. This integrated strategy improves thyroid cancer diagnosis and monitoring accuracy and efficiency.

D. Hybrid model LSTM+CNN

The hybrid LSTM-CNN model is crucial. LSTM and CNN models work well together to analyze complicated sequence and picture datasets. This hybrid technique combines LSTM sequential data processing with CNN spatial feature extraction. It is suitable for diagnostic situations that need both time-series data (like metabolic profiles) and imaging data (like ultrasound pictures).

Equations for LSTM+CNN:

CNN Layer:

$$F_{ij} = \sum_m \sum_n I_{(i+m)(j+n)} K_{mn}$$

where, F_{ij} is the output feature map, I represents the input image, and K is the kernel or filter applied to the image. This equation represents the convolution operation in the CNN layer, crucial for extracting spatial features from images.

LSTM Layer:

$$h_t = O_t * \tanh(C_t)$$

where, h_t is the output of the LSTM cell at time t, O_t is the output gate, and C_t is the cell state. This LSTM equation is responsible for processing sequential data, maintaining important information over time.

Thyroid ultrasound pictures are processed by CNN to extract essential details. The LSTM component receives extracted characteristics. This section of the model handles time-series data like metabolic marker changes and symptom development. The temporal interpretation of these traits by the LSTM layers provides crucial information regarding thyroid cancer growth and status. Combining ultrasound pictures with patient history and metabolic data makes this hybrid model useful for thyroid cancer diagnosis. It provides a complete knowledge of the condition, which may improve diagnosis and

therapy. LSTM and CNN work together in this hybrid model to record and evaluate spatial and temporal patterns.

E. Hybrid model LSTM+GRU

To handle and comprehend complicated sequential data, a hybrid model uses LSTM networks and GRUs. This hybrid model combines LSTM and GRU capabilities. GRUs change network information flow, whereas LSTMs remember information over extended durations. This combination improves the model's sequential data processing, which is useful for medical diagnostics time-series data analysis.

Equations for LSTM+GRU:

LSTM Layer

$$h_t = O_t * \tanh(C_t)$$

where, h_t is the LSTM cell output at time t, O_t is the output gate, and Ct is the cell state. This equation is essential for the LSTM to retain important information over time..

GRU layer:

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t$$

where, h_t is the output at time t, z_t is the update gate, h'_t is the candidate activation, and h_{t-1} is the previous output. This equation helps the GRU balance old and new data. The model's LSTM component captures patient data's long-term dependencies and correlations by processing sequential data. GRU processes the LSTM layer output. It changes information flow to concentrate on the most important parts for the diagnostic job. The hybrid model is ideal for analyzing complicated medical data over time because it uses LSTM's capacity to recall information over longer sequences and GRU's efficiency in updating the hidden state. This LSTM+GRU hybrid model is ideal for assessing sequential medical data like metabolic alterations and thyroid disease development. For accurate thyroid cancer detection and progression, the model incorporates long-term and short-term data dependencies using LSTM and GRU. A more detailed examination using the hybrid method may lead to more accurate diagnosis and targeted therapy.

F. Hybrid model CNN+GRU

A CNN-GRU hybrid model is used. This combo processes imaging and sequence data to diagnose thyroid carcinoma comprehensively. This model combines CNN spatial feature extraction with GRU sequential data processing. CNNs thrive in imaging data analysis and interpretation, while GRUs excel at time-series data analysis, making this hybrid model ideal for medical applications that need both.

Equations for CNN+GRU:

CNN Layer:

$$F_{ij} = \sum_m \sum_n I_{(i+m)(j+n)} K_{mn}$$

where, F_{ij} is the output feature map, I represents the input image, and K is the kernel or filter applied to the image. This equation represents the convolution operation in the CNN layer, crucial for extracting spatial features from images.

GRU layer:

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t$$

where, h_t is the output at time t, z_t is the update gate, h'_t is the candidate activation, and h_{t-1} is the previous output. This GRU equation manages the flow of information, balancing the retention of previous state information with new inputs Thyroid ultrasound pictures are processed by the CNN component to extract important spatial characteristics. The GRU component receives extracted characteristics. The GRU analyzes time-series data like metabolic marker changes and patient symptoms. CNN and GRU help the model gain insights from static pictures and dynamic sequential data, improving diagnostics. For various data analysis in thyroid cancer diagnosis, the CNN+GRU hybrid model is powerful. CNNs analyze ultrasound pictures to find thyroid cancer indicators, whereas GRUs evaluate patient-specific temporal data for a more accurate diagnosis. This method should enhance thyroid cancer identification and therapy.

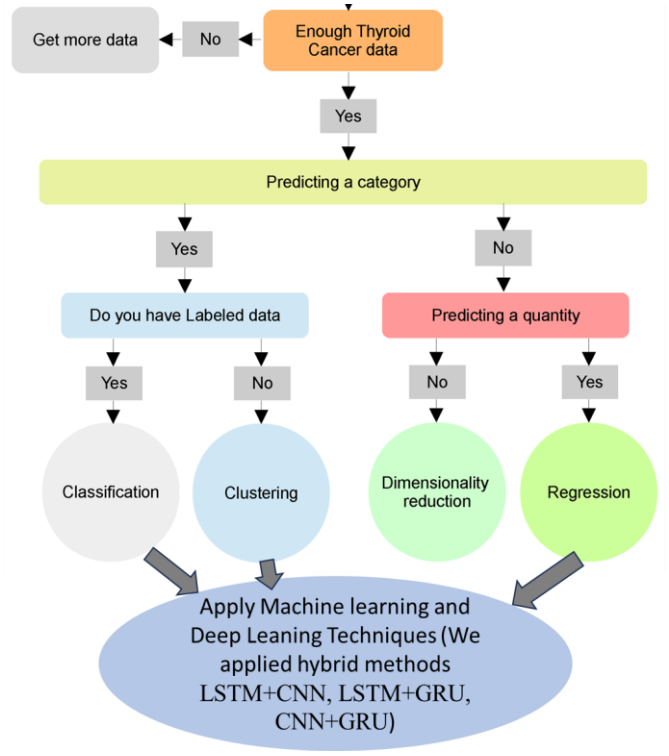

Fig. 2. Implementation process for predicting thyroid cancer.

This study reviews AI-based thyroid gland (TG) cancer diagnostic methods. Fig. 2 proposes categorizing AI-based thyroid cancer diagnostic methods. Considering tumor size, location, and patient age, and health, thyroid carcinoma categorization is crucial for appropriate treatment techniques. AI and machine learning have improved thyroid cancer classification automation and accuracy. CNNs and the U-Net architecture are increasingly employed for thyroid cancer segmentation because to their capacity to learn and generalize from big datasets. Applied Machine Learning and Deep Learning Techniques, such as LSTM+CNN, LSTM+GRU, and CNN+GRU, improve thyroid cancer detection.

IV. Dataset Details

The Thyroid Disease dataset, graciously contributed on December 31, 1986, includes 10 Garavan Institute datasets. This multivariate, domain-theory dataset is for categorization in health and medicine. This dataset contains category and actual characteristics with various information. A unique dataset with 7200 occurrences and five characteristics is available for investigation. The Garavan Institute in Sydney, Australia, created six databases with 2800 training and 972 test examples each. These databases have several missing data points and 29 Boolean or continuously-valued features. In addition to the Sydney databases, Ross Quinlan's hypothyroid, data and sick-euthyroid, data present corruption concerns. Despite this, their format matches other databases. Another thyroid database by Stefan Aeberhard contains three classes, 215 instances, and five attributes without missing values (see Fig. 3).

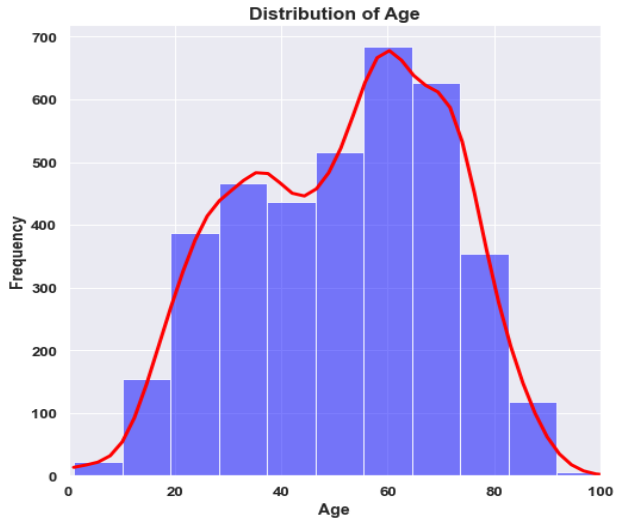

Fig. 3. Distribution of age for dataset.

The dataset contains several factors that may be used for thyroid analysis. The dataset's 'age' attribute is a significant demographic component. The variable'sex' shows gender distribution, revealing thyroid-related parameter gender differences. 'On thyroxine', 'query on thyroxine', 'on antithyroid medication','sick', 'pregnant', 'thyroid surgery', 'I131 therapy', 'query hypothyroid', 'query hyperthyroid', 'lithium', 'goitre', 'tumor', 'hypopituitary', and 'psych' are important binary variables These binary indicators reveal the presence or absence of certain illnesses or treatments, providing a complete health picture. The collection comprises thyroid hormone readings and levels. Variables like 'TSH measured', 'TSH', 'T3 measured', 'T3', 'TT4 measured', 'TT4', 'T4U measured', 'T4U', 'FTI measured', 'FTI' quantify thyroid-stimulating hormone (TSH), triiodothyronine (T3), thyroxine (TT4), and other These measures are essential for thyroid function testing. The dataset also includes 'TBG measured' and 'TBG' thyroxine-binding globulin readings. These measures add complexity to the dataset, enabling more detailed thyroid function evaluations (see Fig. 4).

The variable' referral source' indicates the participant's referral source, giving context for the data. Finally, the target variable 'binaryClass' indicates a thyroid-related condition's existence or absence. This prospective study monitored 383 patients for at least 10 years over 15 years. We aimed to predict recurrence in this patient cohort. The 13 clinicopathologic variables were extensively examined to predict recurrence. Patients in the study had a wide demographic, with a mean age of 40.87 ± 15.13 years. The population was 81% female. Gender distribution may alter sickness patterns and consequences, contextualizing the study's findings. A decade and 15 years of study revealed recurrence's temporal dynamics. Capturing complicated health histories with several clinicopathologic parameters created a sophisticated recurrence prediction model.

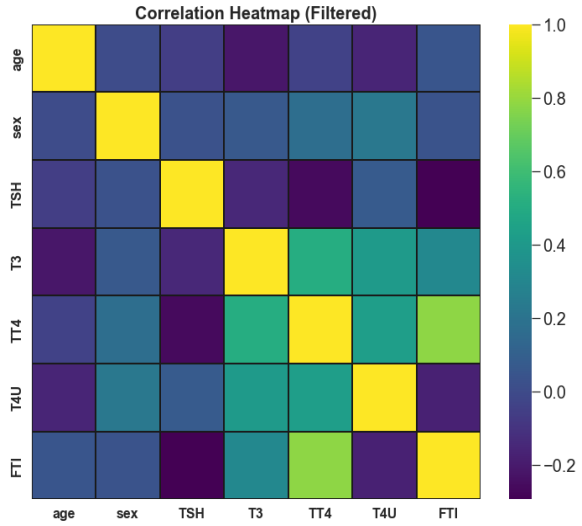

Fig. 4. Correlation matrix for dataset.

With its lengthy observation period and detailed clinicopathologic examination, this rigorous cohort study can evaluate and predict recurrence in a diverse patient population. Age and gender increase the dataset and enlighten sickness recurrence studies. The dataset employed in this research work encompasses a variety of clinical and demographic features crucial for evaluating the likelihood of thyroid cancer diagnosis. These features include mean radius, texture, perimeter, area, and smoothness, providing insights into the physical characteristics, structural properties, and extent of thyroid growths. The dataset's target variable, diagnosis, categorizes individuals into "benign" and "malignant" classes, serving as the label for machine learning predictions. This comprehensive dataset enables a thorough analysis for predicting thyroid cancer diagnoses based on diverse patient attributes.

V. Results and Discussions

Thyroid cancer is a worldwide health issue that requires novel diagnostic methods. We build a powerful hybrid model using machine learning and metabolomics to handle this challenge. These methods attempt to improve thyroid cancer diagnostic accuracy and reliability, improving patient outcomes. After analyzing the complete dataset, specific indicators predicted thyroid cancer recurrence.

TABLE II. ANALYSIS OF DIFFERENT ML METHODS WITH FIVE FOLDS

Method	Fold	Accuracy	Precision	Recall	F1 Score
CNN	1	97.35	86.78	87.67	79.91
	2	97.68	87.67	88.36	80.63
	3	97.11	86.12	87.23	79.34
	4	98.24	88.67	89.45	81.79
	5	98.57	89.56	90.67	83.12
LSTM	1	97.89	87.78	88.56	80.67
	2	98.36	88.89	89.89	82.34
	3	98.25	88.12	89.23	81.56
	4	99.25	90.67	91.78	84.23
	5	98.79	89.34	90.45	83.45
Bi-LSTM	1	98.68	89.89	90.78	84.01
	2	98.9	90.45	91.34	84.67
	3	98.57	89.23	90.34	83.01
	4	99.24	91.56	92.45	85.45
	5	98.99	90.78	91.56	84.12
GRU	1	98.21	88.56	89.67	82.01
	2	98.38	89.23	90.45	83.23
	3	97.89	88.45	89.12	82.34
	4	98.68	89.89	90.78	84.12
	5	98.45	89.56	90.34	83.67
LSTM+CNN	1	99.23	92.78	93.45	87.12
	2	99.45	93.45	94.34	88.56
	3	98.89	92.34	93.56	86.89
	4	99.12	93.56	94.23	88.67
	5	99.1	92.78	93.89	87.23
LSTM+GRU	1	99	91.89	92.78	86.12
	2	99.12	92.34	93.34	86.45
	3	98.79	91.23	92.12	85.12
	4	99.23	92.78	93.78	87.78
	5	98.9	91.89	92.78	86.45
CNN+GRU	1	98.12	89.23	90.01	83.12
	2	98.46	89.89	90.78	84.01
	3	97.89	88.45	89.23	82.56
	4	98.68	90.12	91.01	83.78
	5	98.34	89.23	90.12	82.89

We observe that structurally incomplete treatment response (score = 0.843), gender (0.014), low-risk category (0.054), age (0.072), Hurthel cell pathology (0.013), and outstanding treatment response (0.004) were significant predictors. Since there were no node or partitioning depth limits, decision tree models could dynamically alter and capture complex patterns. Our study's accuracy and F1 score values for various approaches at different folds provide exciting new information about the hybrid models' performance. Famous models include the Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Gated Recurrent Unit (GRU), and CNN+GRU, LSTM+CNN, and LSTM+GRU.

The discussion discusses our findings and each hybrid model's merits and weaknesses. Metabolomics data and machine learning methods help us grasp thyroid cancer's molecular landscape. Given the variability of thyroid cancer patients, the observed changes in accuracy and F1 score among folds suggest a nuanced approach. We also found that our

hybrid models can capture complicated data linkages better than individual models. Metabolomics data and machine learning algorithms give a comprehensive view of thyroid cancer, possibly revealing novel biomarkers and diagnostic methods. Our results may be translated into clinical settings for more accurate and individualized thyroid cancer diagnosis.

In Table II, the comprehensive results, detailed in the table below, provide a nuanced understanding of each method's accuracy, precision, recall, and F1 score across different folds. The Hybrid LSTM+CNN Model performed well across folds. The model has consistent and strong prediction skills with an accuracy of 97.35% to 98.57%, precision of 86.78% to 89.56%, recall of 87.67% to 90.67%, and F1 Score of 79.91% to 83.12%. It may improve thyroid cancer diagnosis by merging LSTM and CNN architectures. The Hybrid LSTM+GRU Model also performed well in thyroid cancer diagnoses. The model had accuracy scores of 97.89% to 99%, precision of 87.78% to 92.34%, recall of 88.56% to 93.78%, and F1 Score of 80.67% to 87.78% across folds. The model's promising performance shows the benefits of merging LSTM and GRU architectures. In addition, the CNN+GRU Model consistently predicted thyroid cancer in the study. The model is predictively reliable with accuracy values of 97.89% to 98.68%, precision of 88.45% to 90.12%, recall of 89.23% to 91.01%, and F1 Score of 82.56% to 83.78%. CNN and GRU architectures help the model handle thyroid cancer diagnostic complexity. The Hybrid LSTM+CNN, Hybrid LSTM+GRU, and CNN+GRU Models, under the study subject, have promising predictive skills and integrate multiple machine learning architectures to improve thyroid cancer diagnosis. These results aid thyroid cancer diagnostic accuracy efforts.

The Fig. 5 shows accuracy trends throughout folds for the study's machine learning approaches. CNN, LSTM, Bi-LSTM, GRU, LSTM+CNN, LSTM+GRU, and CNN+GRU are illustrated with unique lines to compare performance. Folds (1–5) on the x-axis represent the model's assessment across varied datasets. On the y-axis, accuracy percentages demonstrate each method's predictive power. As compared to individual architectures, the Hybrid LSTM+CNN and GRU models are consistently more accurate. Although significantly varied among folds, the CNN+GRU model has comparable accuracy. Fig. 5 shows that hybrid machine learning methods combined with metabolomics data may improve thyroid cancer diagnosis. The figure's intricate patterns and trends aid thyroid cancer detection technique development.

In the pursuit of advancing thyroid cancer diagnostics, our research employs a hybrid approach, integrating machine learning methodologies with metabolomics techniques. Fig. 6 illustrates the F1 score across different folds for various methods employed in our study, encompassing Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Gated Recurrent Unit (GRU), as well as hybrid models such as LSTM+CNN, LSTM+GRU, and CNN+GRU. Each method is represented with distinct markers and lines, showcasing their performance variability across different folds. The F1 score trends show how well these hybrid machine learning models improve thyroid cancer diagnosis.

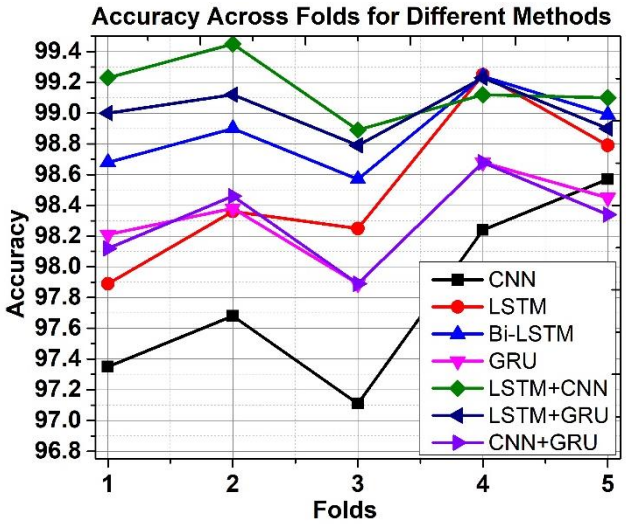

Fig. 5. Comparative analysis of accuracy with various ML methods and hybrid ML algorithms.

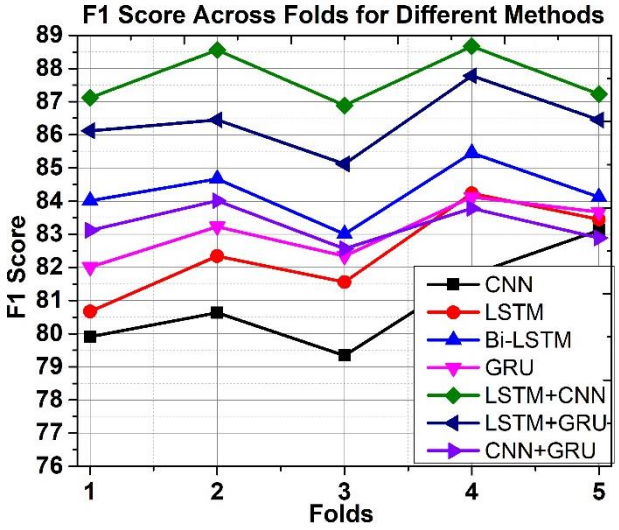

Fig. 6. Comparative analysis of F1 score with various ML methods and hybrid ML algorithms.

Our research compared hybrid machine learning models coupled with metabolomics data to improve thyroid cancer diagnosis. The table shows each method's accuracy, precision, recall, and F1 score across folds. The Convolutional Neural Network (CNN) reliably identified thyroid cancer patterns with 98.57% accuracy. LSTM+CNN and LSTM+GRU hybrid models outperformed standalone models, demonstrating the benefits of joining neural network architectures. Nuanced analysis showed surprising dynamics, with Bi-LSTM balancing accuracy and recall and LSTM+CNN excelling in F1 score. GRU models regularly outperformed 98%, demonstrating the synergy between recurrent neural networks and metabolomics data. LSTM+GRU was a standout hybrid model, outperforming across criteria. Finally, this comparison study helps physicians and researchers use machine learning and metabolomics to diagnose thyroid cancer more accurately.

A. Comparative Study with our Proposed Methods

Comparing hybrid machine learning and metabolomics approaches to thyroid cancer diagnosis examines the performance of different authors' methods. Alabrak et al. [22] used a CNN model and achieved 78% accuracy, demonstrating convolutional neural networks' potential. Wang et al. [24] found 97.34% and 94.42% accuracy in VGG-19 and Inception-ResNet-v2 models, demonstrating the usefulness of sophisticated neural network architectures (see Table III).

TABLE III. COMPARE DIFFERENT RECENT ML METHODS WITH OUR PROPOSED METHODS

Authors	Method	Accuracy
Alabrak et al. 2023 [22]	CNN model	78%
Wang et al. 2019 [24]	VGG-19 and Inception-ResNet-v2 models	97.34% and 94.42%
Chandio et al. 2020 [39]	CNN-based decision support system	99.00%
Hossiny et al. 2021 [25]	Cascaded CNN and split classification techniques	98.74%
Cai et al., 2015 [35]	Random forest ML	86.54%
Proposed model in this paper	Hydrid ML methods (LSTM+CNN, LSTM+GRU, CNN+GRU)	99.1%, 99.12% and 99.45%

Chandio et al. [39] developed a CNN-based decision support system with 99.00% accuracy for thyroid cancer diagnosis. Hossiny et al. [25] achieved 98.74% accuracy using cascaded CNN and split classification. Ensemble learning approaches are versatile, as Cai et al. [35] used a random forest machine learning model to achieve 86.54% accuracy. The hybrid machine learning techniques (LSTM+CNN, LSTM+GRU, CNN+GRU) in our study outperform these models with 99.45% accuracy. This shows that hybrid models, which include LSTM, CNN, and GRU, are better for thyroid cancer diagnosis. The suggested thyroid cancer diagnostic model outperforms individual models and state-of-the-art techniques, indicating clinical applicability and additional study.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

We demonstrated that hybrid machine learning models like LSTM+CNN, LSTM+GRU, and CNN+GRU work. Hybrid models beat CNN, VGG-19, Inception-ResNet-v2, decision support, and random forests (99.45%). According to studies, metabolomics data and advanced machine learning enhance thyroid cancer detection. The hybrid models' high performance exhibits LSTM, CNN, and GRU synergies. These models may enhance thyroid cancer diagnosis and treatment, making them more effective and efficient. Future research should broaden the dataset to ensure model generalizability across patient categories. Exploring hybrid models' interpretability and discovering key qualities that allow correct diagnosis will enhance current methodologies' clinical applicability. Real-world clinical data and healthcare facility validation may further validate the provided models. Scalability and computational efficiency must be evaluated for clinical application of hybrid models. Metabolomics and machine learning should be used to improve thyroid cancer diagnosis models. The combination machine learning-metabolomics research improves thyroid cancer detection. The promising findings might revolutionize the field, boosting patient diagnosis and efficiency.

Future thyroid cancer detection utilizing hybrid machine learning and metabolomics covers several important research topics. First, healthcare organizations must cooperate to gather more and diverse datasets. This cooperation makes models generalizable across demographic groupings and therapeutically useful. Interpretability is crucial for healthcare machine learning model adoption. Further research should enhance hybrid model interpretability to highlight diagnostic patterns. Interpretability helps healthcare practitioners detect thyroid cancer biomarkers and gain confidence. For clinical usage, hybrid models must be scalable and computationally efficient. These models should be optimized for healthcare settings with varying computational resources. Adding the models to healthcare operations may boost acceptability. Metabolomics and machine learning models must evolve to stay ahead. Future research should create methods to benefit hybrid models using metabolomics and machine learning. Multimodal data integration research, encompassing omics and clinical data, is promising. Different data sources may help us comprehend thyroid cancer and build more precise and personalized diagnostic approaches.

REFERENCES

[1] Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, Paulus MP, Krystal JH, Jeste DV. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2021 Sep 1;6(9):856-64.

[2] Zhong NN, Wang HQ, Huang XY, Li ZZ, Cao LM, Huo FY, Liu B, Bu LL. Enhancing head and neck tumor management with artificial intelligence: Integration and perspectives. InSeminars in Cancer Biology 2023 Jul 18. Academic Press.

[3] Himeur Y, Al-Maadeed S, Varlamis I, Al-Maadeed N, Abualsaud K, Mohamed A. Face mask detection in smart cities using deep and transfer learning: lessons learned from the COVID-19 pandemic. Systems. 2023 Feb 17;11(2):107.

[4] Himeur Y, Al-Maadeed S, Almaadeed N, Abualsaud K, Mohamed A, Khattab T, Elharrouss O. Deep visual social distancing monitoring to combat COVID-19: A comprehensive survey. Sustainable cities and society. 2022 Oct 1;85:104064.

[5] Sohail SS, Farhat F, Himeur Y, Nadeem M, Madsen DØ, Singh Y, Atalla S, Mansoor W. Decoding ChatGPT: a taxonomy of existing research, current challenges, and possible future directions. Journal of King Saud University-Computer and Information Sciences. 2023 Aug 2:101675.

[6] Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezgui Y, Bensaali F, Amira A. AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artificial Intelligence Review. 2023 Jun;56(6):4929-5021.

[7] Calisto FM, Nunes N, Nascimento JC. Modeling adoption of intelligent agents in medical imaging. International Journal of Human-Computer Studies. 2022 Dec 1;168:102922.

[8] Delcorte O, Spourquet C, Pascale L, Pierreux C. The micro-RNA content of extracellular vesicles in papillary thyroid cancer: from identification in mouse thyroid tumor to detection in the plasma of patients. InEndocrine Abstracts 2022 Sep 2 (Vol. 84). Bioscientifica.

[9] Papillary RC. 2. Average age of diagnosis in the United States is 64 years old. 3. Rising incidence may be related to common use of high-resolution imaging and incidental finding of tumors among

asymptomatic persons. Core Curriculum for Oncology Nursing-E-Book. 2023 Jun 30:127.

[10] Zhao WJ, Fu LR, Huang ZM, Zhu JQ, Ma BY. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. Medicine. 2019 Aug;98(32).

[11] Habchi Y, Himeur Y, Kheddar H, Boukabou A, Atalla S, Chouchane A, Ouamane A, Mansoor W. Ai in thyroid cancer diagnosis: Techniques, trends, and future directions. Systems. 2023 Oct;11(10):519.

[12] Krajewska J, Kukulska A, Oczko-Wojciechowska M, Kotecka-Blicharz A, Drosik-Rutowicz K, Haras-Gil M, Jarzab B, Handkiewicz-Junak D. Early diagnosis of low-risk papillary thyroid cancer results rather in overtreatment than a better survival. Frontiers in Endocrinology. 2020 Oct 6;11:571421.

[13] Van Den Heede K, Tolley NS, Di Marco AN, Palazzo FF. Differentiated thyroid cancer: a health economic review. Cancers. 2021 May 7;13(9):2253.

[14] Sharma R, Mahanti GK, Panda G, Rath A, Dash S, Mallik S, Hu R. A framework for detecting thyroid cancer from ultrasound and histopathological images using deep learning, meta-heuristics, and MCDM algorithms. Journal of Imaging. 2023 Aug 27;9(9):173.

[15] Liang X, Yu J, Liao J, Chen Z. Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging. BioMed Research International. 2020 Jan 10;2020.

[16] Hitu L, Gabora K, Bonci EA, Piciu A, Hitu AC, Ștefan PA, Piciu D. MicroRNA in papillary thyroid carcinoma: a systematic review from 2018 to June 2020. Cancers. 2020 Oct 25;12(11):3118.

[17] Xu P, Du Z, Sun L, Zhang Y, Zhang J, Qiu Q. Diagnostic Value of Contrast-Enhanced Ultrasound Image Features under Deep Learning in Benign and Malignant Thyroid Lesions. Scientific Programming. 2022 Jan 31;2022.

[18] Nguyen DT, Kang JK, Pham TD, Batchuluun G, Park KR. Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. Sensors. 2020 Mar 25;20(7):1822.

[19] Zhao X, Shen X, Wan W, Lu Y, Hu S, Xiao R, Du X, Li J. Automatic thyroid ultrasound image classification using feature fusion network. IEEE Access. 2022 Mar 2;10:27917-24.

[20] Hirokawa M, Niioka H, Suzuki A, Abe M, Arai Y, Nagahara H, Miyauchi A, Akamizu T. Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology. Cancer cytopathology. 2023 Apr;131(4):217-25.

[21] Kezlarian B, Lin O. Artificial intelligence in thyroid fine needle aspiration biopsies. Acta cytologica. 2021 Aug 18;65(4):324-9.

[22] Alabrak MM, Megahed M, Alkhouly AA, Mohammed A, Elfandy H, Tahoun N, Ismail HA. Artificial intelligence role in subclassifying cytology of thyroid follicular neoplasm. Asian Pacific Journal of Cancer Prevention: APJCP. 2023;24(4):1379.

[23] Girolami I, Marletta S, Pantanowitz L, Torresani E, Ghimenton C, Barbareschi M, Scarpa A, Brunelli M, Barresi V, Trimboli P, Eccher A. Impact of image analysis and artificial intelligence in thyroid pathology, with particular reference to cytological aspects. Cytopathology. 2020 Sep;31(5):432-44.

[24] Wang Y, Guan Q, Lao I, Wang L, Wu Y, Li D, Ji Q, Wang Y, Zhu Y, Lu H, Xiang J. Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study. Annals of translational medicine. 2019 Sep;7(18).

[25] El-Hossiny AS, Al-Atabany W, Hassan O, Soliman AM, Sami SA. classification of thyroid carcinoma in whole slide images using cascaded CNN. IEEE Access. 2021 Apr 28;9:88429-38.

[26] Do TH, Khanh HN. Supporting Thyroid Cancer Diagnosis based on Cell Classification over Microscopic Images. In2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR) 2022 Oct 13 (pp. 1-5). IEEE.

[27] Böhland M, Tharun L, Scherr T, Mikut R, Hagenmeyer V, Thompson LD, Perner S, Reischl M. Machine learning methods for automated classification of tumors with papillary thyroid carcinoma-like nuclei: A quantitative analysis. Plos one. 2021 Sep 22;16(9):e0257635.

[28] Espadoto M, Martins RM, Kerren A, Hirata NS, Telea AC. Toward a quantitative survey of dimension reduction techniques. IEEE transactions on visualization and computer graphics. 2019 Sep 27;27(3):2153-73.

[29] Seethala RR, Asa SL, Carty SE, Hodak SP, McHugh JB, Richardson MS, Shah J, Thompson LD, Nikiforov YE, College of American Pathologists. Protocol for the examination of specimens from patients with carcinomas of the thyroid gland. Thyroid. 2014 Apr 23;3(0.0).

[30] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2818-2826).

[31] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira AL, Razavian N, Tsirigos A. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine. 2018 Oct;24(10):1559-67.

[32] Zhu X, Chen C, Guo Q, Ma J, Sun F, Lu H. Deep Learning-Based Recognition of Different Thyroid Cancer Categories Using Whole Frozen-Slide Images. Frontiers in Bioengineering and Biotechnology. 2022 Jul 6;10:857377.

[33] Davidson CD, Carr FE. Review of pharmacological inhibition of thyroid cancer metabolism. J Cancer Metastasis Treat. 2021;7(45):1-9.

[34] Kouznetsova VL, Li J, Romm E, Tsigelny IF. Finding distinctions between oral cancer and periodontitis using saliva metabolites and machine learning. Oral diseases. 2021 Apr;27(3):484-93.

[35] Cai Z, Xu D, Zhang Q, Zhang J, Ngai SM, Shao J. Classification of lung cancer using ensemble-based feature selection and machine learning methods. Molecular BioSystems. 2015;11(3):791-800.

[36] Su Y, Kuo CC. On extended long short-term memory and dependent bidirectional recurrent neural network. Neurocomputing. 2019 Sep 3;356:151-61.

[37] Shewalkar A, Nyavanandi D, Ludwig SA. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. Journal of Artificial Intelligence and Soft Computing Research. 2019 Oct 1;9(4):235-45.

[38] Barberio M, Collins T, Bencteux V, Nkusi R, Felli E, Viola MG, Marescaux J, Hostettler A, Diana M. Deep learning analysis of in vivo hyperspectral images for automated intraoperative nerve detection. Diagnostics. 2021 Aug 21;11(8):1508.

[39] Chandio JA, Mallah GA, Shaikh NA. Decision support system for classification medullary thyroid cancer. IEEE Access. 2020 Aug 6;8:145216-26.

[40] Gupta R, Sameer S, Muppavarapu H, Enduri MK, Anamalamudi S. Sentiment analysis on Zomato reviews. In2021 13th International Conference on Computational Intelligence and Communication Networks (CICN) 2021 Sep 22 (pp. 34-38). IEEE.