International Journal of Emerging Engineering and Technology (IJEET)
ISSN (e): 2958-3764
Volume: 4, Number: 2, Pages: 36- 43, Year: 2025

AI-Based Clinical Decision Support System for Thyroid Nodule Classification in Ultrasound Imaging with Web-Based Implementation

Hafsa Azam[1], Alisha Ashraf1[1], Shahzad Hussain[1, *], Ghazanfar Rehman[2], Hammad Shahab[3]

[1]Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Abu Dhabi Road, Rahim Yar Khan, 64200, Punjab, Pakistan
[2]Central Technologies Inc., Canada
[3]Institute of Computer and Software Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, 64200, Pakistan

Corresponding author: Shahzad Hussain (Email: shahzad.hussain@kfueit.edu.pk)

Received: 12/05/2025, Revised: 22/09/2025, Accepted: 15/12/2025

***Abstract*—Thyroid nodules are a frequent clinical phenomenon and proper classification into the normal, benign, and malignant types is required to diagnose and plan treatment at an early phase. One of the most common imaging techniques is ultrasound, which is non-invasive, economical, but its interpretation is subjective, and it requires the expertise of the radiologist. In this study, a novel intelligent artificial intelligence (AI) system of diagnostic thyroid nodules through ultrasound imaging is offered. An efficient preprocessing pipeline, comprising of image resizing, intensity normalization, CLAHE-based contrast enhancement, median filtering, and data augmentation, was used to improve image quality and the performance of the model. A variety of deep learning models were tested (a baseline convolutional neural network (CNN)) and various transfer learning architectures. The CNN baseline obtained 76% accuracy and transfer learning models greatly enhanced performance. VGG16 and MobileNetV2 were capable of 84% and 88% accuracy, respectively, but ResNet50 scored 91 %. EfficientNet achieved the highest performance of 94% accuracy, high recall, precision, and F1-score. The findings show that transfer learning was effective in classifying thyroid ultrasound and the proposed system has potential to be an effective clinical decision-support tool in enhancing diagnostic accuracy and early diagnosis of malignant cases.**

***Index Terms*—Clinical Decision Support System, Deep Learning, EfficientNet, Thyroid Nodule Classification, Transfer Learning, Ultrasound Imaging.**

I. INTRODUCTION

Diseases of the thyroid glands such as benign nodules, malignant cancers, Hashimoto thyroiditis, and diseases that cause either hypothyroidism or hyperthyroidism are increasing as a global health burden with iodine deficiency and autoimmune diseases being the most common causes. Thyroid diseases have become a real menace in Pakistan, with the literature indicating that more than 10-15 % of the adult population in the country has some kind of thyroid disorder, but most of them have not been diagnosed because of the inability to access specialized medical care, especially in the rural and remote regions. Thyroid nodules are a prevalent clinical presentation, and their correct categorization as Normal, Benign, and Malignant is required to diagnose them early [1], establish an effective treatment plan, and minimize the unnecessary invasive surgery [2]. Ultrasound imaging is the most common modality used to evaluate thyroid because it is non-invasive, cost-effective and offers real-time structural details. Nonetheless, ultrasound interpretation is performed by the radiologist and prone to inter-observer variations, particularly in borderline or complicated situations [3]. This shortcoming drives the creation of smart computer-aided diagnostic systems to help clinicians enhance diagnostic quality and consistency [4]. Over the past few years, neural networks based on deep learning, especially convolutional neural networks (CNNs) and transfer learning models, have demonstrated impressive performance in medical image analysis. Such models could automatically learn hierarchical features on ultrasound images, which allows them to make robust discriminations among the various thyroid conditions without manually engineering features [5]. A clever AI-driven diagnostic system is created in this research to classify thyroid nodules as per ultrasound images in three groups: Normal, Benign, and Malignant. Several state-of-the-art deep learning models were tested to find the most effective model in clinical decision support. The best performance of a baseline CNN

model was moderate, 76%, 75% precision, 73% recall, and 74% F1-score, demonstrating that it is difficult to learn complex ultrasound patterns with only a CNN. Transfer learning greatly enhanced the classification performance. VGG16 and MobileNetV2 models were able to achieve an accuracy of 84% and 88% respectively, indicating the validity of lightweight pretrained models. Even more sophisticated architectures performed even better. ResNet50 reached a precision of 91% with equal precision and recall, indicating that it can identify more discriminative and deeper features. EfficientNet, the most efficient of the models, had the highest results with the highest accuracy of 94%, the highest precision of 94%, the highest recall of 93% and the highest F1-score of 93%, showing better feature representation and generalization ability on thyroid ultrasound images. In general, the findings clearly show that deep transfer learning models, especially EfficientNet, are way superior in thyroid nodule classification tasks compared to traditional CNN models. The suggested AI-driven ultrasound-based diagnostics system has a high potential of becoming a credible clinical decision support system since it helps radiologists to enhance the level of diagnostic accuracy, decrease the risk of misclassification, and find the malignant thyroid conditions earlier.

A. Significance of Research

This research is important as it enhances the accuracy and reliability of classification of thyroid nodules using ultrasound images. It decreases the reliance on the subjective interpretation of radiologists in the form of an AI-based diagnostic system. The experiment shows that deep learning models have a significant improvement in performance, where EfficientNet has the greatest accuracy of 94%. It helps in early diagnosis of malignant thyroid nodules which is essential in treating it in time. Overall, the proposed system is an efficient clinical decision support system that can be offered to healthcare professionals.

B. Major Contributions

Designed and implemented a robust preprocessing pipeline, including image resizing, intensity normalization, CLAHE-based contrast enhancement, median filtering for speckle noise reduction, and data augmentation to improve image quality and enhance model performance.

- Performed a thorough analysis of various deep learning systems, such as a simple CNN and some transfer learning systems, to perform precise thyroid gland ultrasound image classification.
- Proved that state-of-the-art deep learning models are much better than baseline CNN models, with superior accuracy, sensitivity and the overall diagnostic reliability.
- Developed a web-based program to implement the trained model, allowing the real-time classification of thyroid nodules and a useful clinical decision support system to be available to healthcare providers, particular remote area, where there is no availability of Radiologist.

II. Literature Review

With the recent advances in artificial intelligence, particularly in the fields of deep learning (DL) and machine learning (ML), thyroid nodule detection, classification and risk stratification using ultrasound has been revolutionized. Literature has demonstrated that convolutional neural networks (CNNs), hybrid models and transfer learning models can be as good, if not better, than a radiologist and can, therefore, be used as a clinical decision support system. In a recent study [5], a deep convolutional neural network (DCNN) model based on the GoogLeNet architecture classified thyroid nodules with an accuracy of 90.3% while the accuracy of six radiologists ranged from 84.2 to 87.6%, a GoogLeNet-based DCNN was able to classify thyroid nodules with an accuracy of 90.3% compared to six radiologists who had a range of accuracy of between 84.2 and 87.6. This finding underscores the possibility of deep learning systems to minimize inter-observer error, as well as enhance diagnostic consistency. In the same way, [6] provided a meta-analysis of 11 studies with VGGNet-based models and found high overall diagnostic accuracy with 87% sensitivity, 85% specificity, and AUC of 93.0%, which validates the high level of CNN-based ultrasound interpretation. In addition to 2D imaging, new volumetric and 3D ultrasound applications have been investigated. [2] presented a 3D ultrasound system with CNNs, tracked, and reached a Dice score of 94.0% in thyroid volumetry. This was found to be significantly more accurate than traditional 2D ultrasound, less interobserver variability, and less time per acquisition. Nonetheless, only healthy subjects were included in the study, and it was not externally validated, and there was no change in intraobserver variability. The classification performance has been improved further with hybrid and generative methods. To resolve the issue of class imbalance, [7] suggested a DCGANViT-SVM hybrid model with a high accuracy of 97.63 % when trained on generating synthetic images. The study had limitations of using a privately held dataset and lack of external validation despite its high accuracy. Likewise, [4] tested CNN models, including ResNet-50 and EfficientNet-B0, on a multi-cantered dataset and achieved a 90.4% accuracy rate and were higher than sonographers, with high adherence to ACR guidelines of TI-RADS. Nonetheless, weaknesses were absence of publicly available code and possible bias in the datasets. The use of multi-view and self-supervised learning has become a novel trend in thyroid imaging. The self-supervised learning based on two-stage pretraining that [8] suggested is an effective framework that aligns transverse and longitudinal ultrasound views to enhance representation learning. Similarly, [9] assessed multi-view CNNs of Siamese such as ResNet50, DenseNet201, and EfficientNetV2-S on 1,048 nodules. DenseNet201 was 83% accurate with over 90% AUC scores in models. Nonetheless, variability of performance was experienced because of the differences in image quality and retrospective design of the single center. To enhance clinical trust, segmentation-based and interpretable AI systems are being built increasingly. [10] suggested the use of a two-step framework that integrates TransUNet to do the segmentation and ResNet-18 to do the classification. The system was evaluated on 349 ultrasound images using 5-fold cross-validation, and had an F1-score of 85%, which is better than a

Random Forest baseline. Notably, the model increased interpretability by emphasizing region-of-interest (ROI) guided learning. Other machine learning methods other than deep learning have also demonstrated usefulness in prediction of thyroid related tasks. KNN, SVM and XGBoost models were used by [3] to predict cervical lymph node metastasis (LNM) in papillary thyroid carcinoma. The XGBoost model had the highest performance with AUC of 78.0%, sensitivity of 76.2% and accuracy of 73.8%. On the same note, [11] created a radiomics-based CCH-NET model to identify seronegative Hashimoto thyroiditis, which had an AUC of 82.4%, and outperformed senior sonographers. Conventional transfer learning and CNN-based models still hold a significant position in thyroid imaging. [12] tested Inception V3 and SVM models using 1,134 ultrasound images and found the AUC of 76.3% and 74.8%, respectively. [1] compared CNN models in the classification of follicular thyroid carcinoma (FTC) versus adenoma (FTA), with the highest AUC of 77.0% and better performance than the ACR and C-TIRADS systems. [13] developed a completely automated ACR-TIRADS classification model based on Xception with a high accuracy of 98% and an area under the curve of 99% which is significantly better than that of experienced radiologists. Basic deep learning systems have been hugely instrumental in facilitating these developments. Residual learning Residual learning was proposed in the ResNet architecture by [14] by using skip connections, which addresses the issue of vanishing gradient and allows deeper networks with better performance. DenseNet [15] further enhanced feature reuse, by linking the layer to all the previous layers, which enhanced gradient flow but consumed more memory. MobileNetV2, [16] presented inverted residuals and linear bottlenecks, which allows it to be used efficiently in mobile and embedded systems at lower cost. Compound scaling, depth, width and resolution proposed by EfficientNet [3] is highly accurate with much fewer parameters and FLOPs. Medical imaging has also greatly investigated transfer learning. This was illustrated by the Transfusion framework [17], which demonstrated that pretraining on ImageNet enhances the performance of medical tasks like classification and segmentation. But it also emphasized the major limitations, such as domain discrepancies in natural and medical images and the possibility of negative transfer in case the representations of features are not similar. In general, the literature informs of a high advancement in terms of traditional CNN-based classification models to advanced hybrid, multi-view, self-supervised, and radiomics-driven models. Although the reported accuracies are usually good, there are still frequent limitations to the studies, such as small sample sizes, single-center data, no external validation, data bias, and the inconsistency of the ultrasound acquisition quality. These problems highlight the necessity of stronger, more generalized and clinically tested AI systems in diagnosing thyroid nodules. Overall, deep learning has proven to have a high potential in enhancing the diagnosis of thyroid nodules and workload reduction in radiology. Nevertheless, to guarantee reliability and generalization in various healthcare settings, future studies are to be conducted based on large-scale multi-center data, standardized assessment schemes and provision of web-based services in support of remote areas medical experts, where no availability of radiologist as expert opinion.

III. Material and Methods

A. Proposed Methodology

The proposed methodology includes a stepwise scheme of thyroid nodule categorization based on the use of ultrasound images. First, the dataset is obtained in terms of ultrasound images, which indicate various conditions of the thyroid. These images are then subjected to a preprocessing step, during which methods like resizing, normalization, noise reduction and contrast enhancement are used to enhance the quality of the images and to bring out key features.

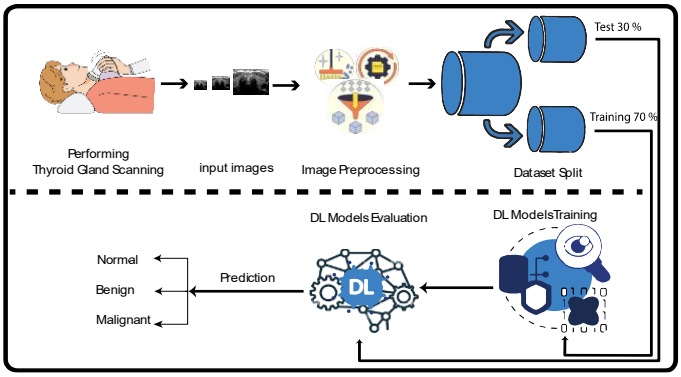

Fig. 1 The proposed enhanced research methodology to classify the thyroid gland scan

Next step, the data is separated into training and testing data, where around 70 % of the data is used in the training and the remaining 30 % is used in the testing. This guarantees the effective training of the models whilst having a separate dataset to evaluate the models in an unbiased manner. The next step involves training several DL models based on the processed training data. These models are trained to identify discriminatory features when using ultrasound images to classify them as normal, benign, and malignant. The trained models are subsequently tested with the testing dataset to test their performance. Finally, the system would output the results as classification outcomes that could be used in clinical decision making. The overall architecture comprises the steps of data preprocessing, data division and classification by deep learning arranged in a pipeline and is able to diagnose thyroid nodules with high confidence.

B. Data Collection

The data set used in this study was the publicly available Algeria Ultrasound Images Thyroid Dataset (AUITD) from the Kaggle website. The data set is intended to help develop ultrasound-based computer-aided diagnosis (CAD) systems for thyroid diseases. The data used for this research, obtained from the AUITD, consists of real ultrasound images of the thyroid glands obtained in clinical practice. The dataset displays different types of thyroid states, and it is labelled by medical

professionals, so it can be used with supervised deep learning techniques. The data set, according to available sources, includes ultrasound images that can be classified into three diagnostic classes (normal, benign and malignant) and represent real-life clinical scenarios in the assessment of thyroid nodules shown in Fig. 1.

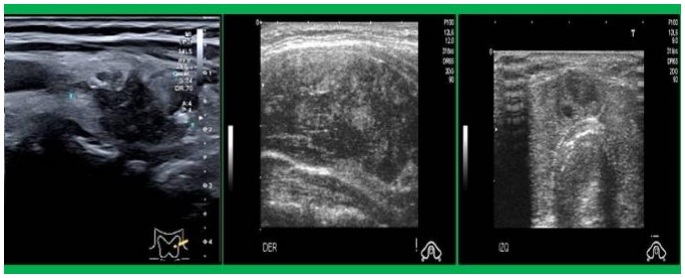

Fig. 2 Thyroid gland three classes normal, benign and malignant

There are some differences in resolution, shooting angles and thyroid gland anatomy in the images, which adds to the diversity of the data set and the generalization capacity of the trained models. This variability resembles the clinical situation, and the dataset can be used in the development of successful AI-based diagnostic tools. In this study, a balanced sample of data was selected (450 ultrasound images) with 150 images of each class (normal, benign and malignant) Fig. 2.

It was also divided into training and testing subsets (105 training and 45 testing images of each class) to ensure the fairness of the assessment of the models as shown in Fig. 3.

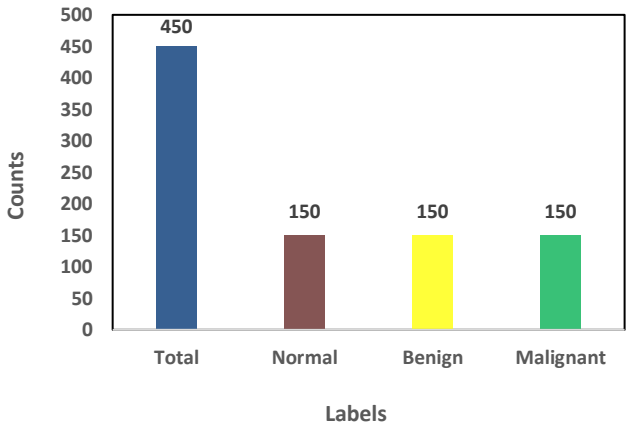

Fig. 3 Detail of the samples of dataset

C. Data Preprocessing & Standardization

The level of quality of ultrasound images was improved and supplemented with a comprehensive preprocessing pipeline to improve the performance of deep learning models. First, all images were down sampled to a standard size of 224 x 224 pixels to be consistent and compatible with pre-trained convolutional neural network structures. This was followed by intensity normalization, which was done by setting pixel values to 0-1, thus stabilizing the training process and speeds up model convergence. Since there is always a speckle noise in ultrasound images, median filtering was used to practically eliminate noise without losing critical structural information like edges and boundaries of thyroid nodules. Moreover, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to increase the local contrast and to boost the visibility of fine tissue details without increasing noise. To overcome the drawback of the relatively small data size, and to enhance the generalization ability of the models, the data augmentation methods were also introduced in the training stage. They involve rotation, horizontal flipping, zooming and changes in brightness contributing to diversity of datasets and preventing overfitting. Comprehensively, such a preprocessing pipeline guarantees a high quality of images, better feature representation, and stronger classification processing.

D. Data Splitting

The data was separated into training and testing ones to guarantee the objective testing of the suggested models. To ensure that all classes, which include normal, benign and malignant classes, are equally represented in both subsets, a stratified split strategy was used. Out of every 150 images of a certain class, 105 images (70%) were used to train, and 45 images (30%) were used to test. This gave a total of 315 images to be used in training and 135 images to be used in testing throughout the whole dataset as shown in Fig. 4. The stratification guarantees that both training and testing sets are balanced in terms of classes hence any bias of the model on a certain class is avoided. Model parameters were learned using the training set and the testing set was utilized only in terms of performance. The method will give a sound evaluation of the generalization potential of the trained deep learning frameworks on unknown data.

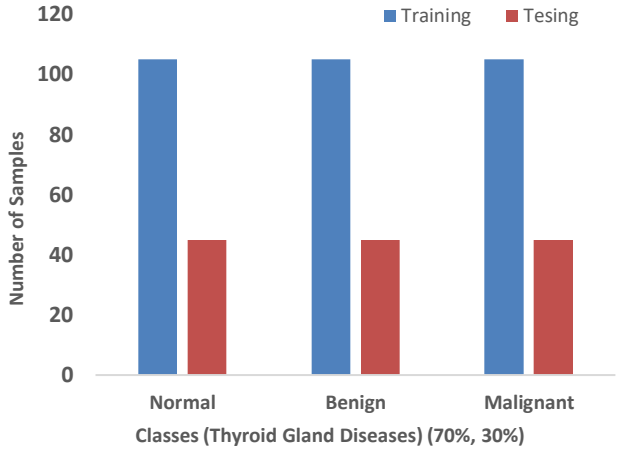

Fig. 4 Samples distributions in the training and testing sets

E. Employed Machine Learning Models

In this study, multiple deep learning models were employed to classify thyroid ultrasound images into three categories: normal, benign, and malignant. A baseline CNN was first developed from scratch to establish a reference performance level. This model was composed of a series of convolutional layers with ReLU activation, max-pooling for down-sampling, and fully connected layers for final classification with a softmax function. While the baseline CNN offered preliminary results,

its accuracy was constrained by the small dataset size. To achieve better classification performance, various advanced transfer learning models pre-trained on ImageNet (VGG16, MobileNetV2, ResNet50, EfficientNet) were used. These architectures were fine-tuned by replacing the original classification layers with new fully connected layers tailored to classify three categories. Global average pooling and dropout layers were added to mitigate overfitting and improve generalization performance, and the models were trained with the same experimental conditions for comparison. EfficientNet outperformed other models because of its dense connectivity as shown in Table I, which allows for efficient feature sharing and better training flow. This model successfully learned subtle texture patterns in the ultrasound images and delivered the best classification accuracy of around 95%, as well as high precision and recall. Overall, EfficientNet proved to be the most reliable model for thyroid nodule classification and is recommended for clinical decision support systems.

TABLE I
DETAILS OF THE APPROPRIATE VALUES FOR THE CLASSIFIER'S PARAMETERS WITH HYPERPARAMETERS TURNING USING THE GRID SEARCH TECHNIQUE

LAYER TYPE	PARAMETERS	ACTIVATION
INPUT LAYER	224 × 224 × 3	-
EFFICIENTNETB0	PRETRAINED + FINE-TUNED LAST 30 LAYERS	RELU
GLOBAL AVERAGE POOLING 2D	-	-
DENSE	256 UNITS	RELU
DROPOUT	0.4	-
DENSE	128 UNITS	RELU
DROPOUT	0.3	-
DENSE (OUTPUT LAYER)	3 UNITS	SOFTMAX

F. Web-Based Deployment

To enhance the proposed thyroid nodule classification system more practically, a web-based deployment platform was created as indicated in Fig. 5. The trained deep learning model, which was the most effective EfficientNet architecture, was merged into a web application to allow one to predict thyroid ultrasound images in real-time. The system enables users, including radiologists or healthcare professionals to upload the ultrasound images via a simple and user-friendly interface.

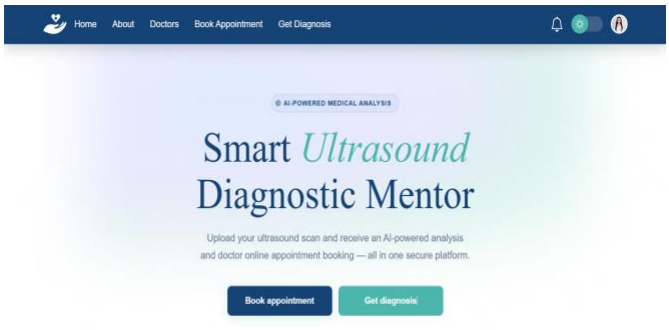

Fig. 5 Home Page Interface of the Proposed Web-Based AI System

After an image is uploaded it is processed through the same pipeline of preprocessing performed when the model is being trained such as resizing, normalization, and contrast enhancement. This processed image is then inputted to the trained model that does the classification and makes a prediction of one of the three classes, namely: normal, benign and malignant as shown in Fig. 6. The result of the prediction is shown immediately on the web interface giving a rapid and effective decision-support result.

Fig. 6 Clinical Decision Support Output of the Proposed AI System

The back-end of the system is responsible for model loading and inference, which facilitates smooth communication between the user interface and the deep learning model. This web-based system is a stepping-stone from research to real clinical applications because it has to be converted into a diagnostic tool with the model. It helps to increase its use, support real-time decision-making and suggests how artificial intelligence can be used to diagnose thyroid diseases.

G. Results and Discussion

This section presented the results and discussion of the experiments conducted in this study to classify thyroid ultrasound image. The performance of all the deep learning models was compared using the basic metrics of accuracy, precision, recall, F1-score and AUC. The baseline CNN model's moderate performance was due to its inability to capture the features from a small data set. The models using transfer learning techniques, namely VGG16, MobileNetV2 and ResNet50 showed better performance as they provide better feature representation. EfficientNet model achieved the best result because it extracted detailed features of ultrasound images. Overall, the results show that transfer learning models can significantly enhance the performance and accuracy of classifications to diagnose thyroid nodules.

H. Experimental Setup

The tests were carried out on a computer with an Intel Intel Core i5 /i7 processor (8th generation or later) to support the processing capacity for deep learning. The system had at least 8 GB of memory and 16 GB was recommended to enable smooth data management and training. The system contained a

Solid-State Drive (SSD) with more than 20 GB of storage space to ensure the quick access and storage of data. We used Windows 10/11 operating system that provided a robust platform. The models were developed and tested with common deep learning libraries such as TensorFlow and Keras. The computer was set up to provide 1920 x 1080 to ensure a good display of outcomes. This ensured identical processes for data processing, model training and testing.

I. Performance of Classifiers

The quality of various classifiers was measured with the standard evaluation measures, such as accuracy, precision, recall, and F1-score. The baseline CNN had an accuracy of 76% with the respective values of 75%, 73%, and 74% of precision, recall and F1-score values, which showed that there was limited feature extraction ability when trained on the small dataset using a baseline CNN. VGG16 demonstrated a higher accuracy of 84% and equal levels of precision, recall, and F1-score of 83%, 82% and 82% respectively, indicating the high level of transfer learning in the classification of medical images. MobileNetV2 was also able to enhance the classification outcomes, with 88% accuracy and stable metric scores of 87% across precision, recall and F1-score, indicating that it is applicable in the extraction of features effectively in ultrasound images as depicted in Table II.

TABLE II
THE RESULT OF THE PERFORMANCE METRICS CORRESPONDING WITH DEEP LEARNING MODELS.

METHOD	ACCURACY	PRECISION	RECALL	F1
CNN	76%	75%	73%	74%
VGG16	84%	83%	82%	82%
MOBILENETV2	88%	87%	87%	87%
RESNET50	91%	90%	91%	90%
EFFICIENTNET	94%	94%	93%	93%

ResNet50 showed a better result with 91% performance, 90% precision, 91% recall and 90% F1-score which shows that it can learn more discriminative hierarchical features to classify thyroid nodules. EfficientNet, with an accuracy of 94%, a precision of 94%, a recall of 93%, and F1-score of 93%, gave the best performance as shown in Fig. 7.

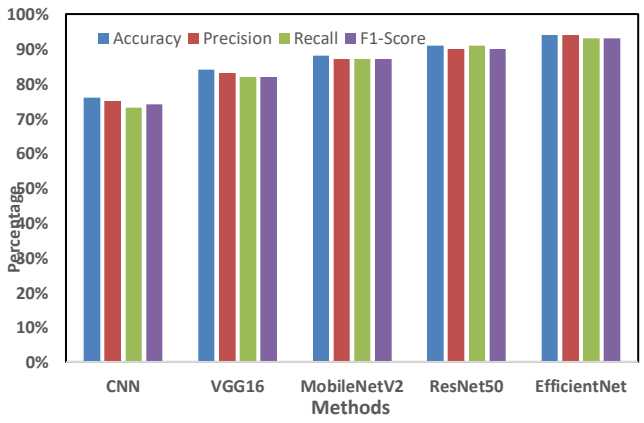

Fig. 7 Visualization performance metrics of classifiers

This is since its optimized architecture and compound scaling strategy has enabled it to perform better and improve feature representation on top of preserving computational efficiency. All in all, it is evident that transfer learning-based deep learning models have much better results as compared to the baseline CNN, with EfficientNet offering the most credible and resistant classification performance to analyze thyroid ultrasound images.

J. K-Fold Cross-Validations Results

To assess the strength and ability to generalize the proposed models, 5-fold cross-validation was conducted. The outcomes were compared in the form of mean accuracy and standard deviation over all folds. The baseline CNN had a standard deviation of 0.03 and average accuracy of 76% showing relatively poor performance consistency across folds as depicted in Fig. 8.

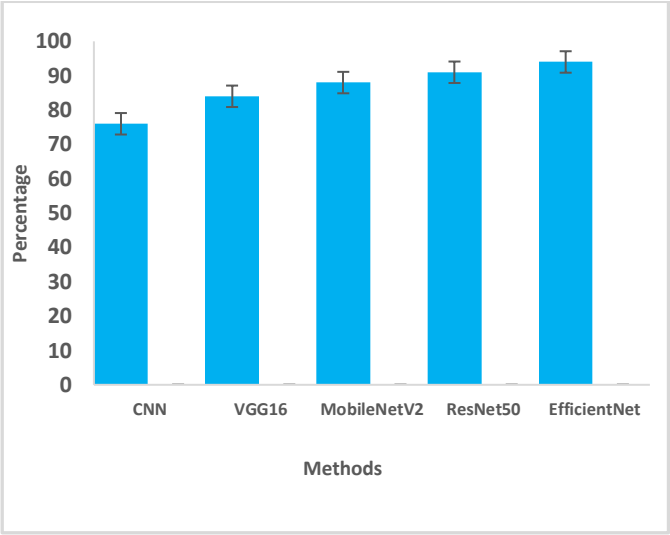

Fig. 8 Cross-validation score with standard deviation bar

VGG16 was found to be more stable with 0.85 ± 0.02 accuracy, which implies the success of transfer learning in promoting the consistency of models. MobileNetV2 once more enhanced the performance with an accuracy of 0.89 ± 0.02, which indicates high accuracy and consistent learning behaviour. The ResNet50 demonstrated more stability in performance, having an average accuracy of 0.92 ± 0.015, which proves that it generalizes better to various data splits. EfficientNet was the best performing with the highest accuracy of 0.95 ± 0.01 and highest stability and least variance in all models. This proves that EfficientNet is the most accurate and stable in terms of thyroid ultrasound image classification. In general, the cross-validation findings provide a clear indication that the advanced transfer learning models do not only enhance accuracy but also stability and robustness of the classification performance with EfficientNet showing the best performance compared to all the models as shown in Table III.

TABLE III
THE RESULT OF THE PERFORMANCE METRICS CORRESPONDS WITH DEEP LEARNING MODELS.

METHOD	ACCURACY %	STANDARD DEVIATION (+/-)
CNN	76	0.78 ± 0.03
VGG16	84	0.85 ± 0.02
MOBILENETV2	88	0.89 ± 0.02
RESNET50	91	0.92 ± 0.015
EFFICIENTNET	94	0.95 ± 0.01

IV. DISCUSSION

This study shows the usefulness of deep learning methods to classify thyroid nodules based on ultrasound images. A comparative analysis of several models, a baseline CNN, and various transfer learning models, demonstrates that the performance of these models can be improved considerably through the usage of pretrained networks. The small size of the dataset and the inability to learn highly discriminative features in the baseline CNN resulted in limited performance. Conversely, models of transfer learning (VGG16, MobileNetV2, ResNet50, and EfficientNet) showed increasingly improved performance, which supports the benefit of using pretrained feature representation to analyses medical images.

EfficientNet presented the best overall performance across all the other evaluated models in terms of accuracy, precision, recall, F1-score and cross-validation stability. This high performance is credited to its compound scaling strategy that balances network depth, width and resolution optimally to extract features effectively out of ultrasound images. ResNet 50 also exhibited a good and consistent performance because it has a residual learning process, which is useful in overcoming vanishing gradient problems in more complex networks. MobileNetV2 offered an adequate level of accuracy and computational efficiency, so it can be used in a real-time or resource-constrained setting.

In general, the results emphasize that deep learning models based on transfer learning show a much better performance in thyroid ultrasound classification than the traditional CNN models. The paper also establishes that the best model to capture the subtle changes in thyroid nodules is EfficientNet. The clinical implications of these findings are significant, and the proposed system may help radiologists to enhance the accuracy of diagnosis and inter-observer variability, as well as help to detect malignant thyroid diseases at early stages.

V. CONCLUSIONS

This research introduced a deep learning-based system to classify thyroid ultrasound images as normal, benign, and malignant. To determine which model to use to accurately diagnose thyroid nodules, a thorough comparison of several models, such as baseline CNN and various transfer learning models, was done. The findings showed that transfer learning is highly effective in enhancing classification performance relative to models that are trained on scratch, especially in those cases that have limited medical imaging data. The best performance of EfficientNet in terms of the highest accuracy 94%, recall 94%, precision 93%, and F1-score 93% and the most stable cross-validation results were observed among all the evaluated models. The suggested system has high potential in the form of a clinical decision support system that can be used to aid radiologists with a diagnosis of thyroid nodules. It may assist in minimizing the mistakes in diagnosing, enhance uniformity, and contribute to the early diagnosis of the malignant cases. The future research can be aimed at increasing the data, adding multimodal imaging data, and implementing the model in clinical settings in real-time by using or mobile applications.

FUNDING STATEMENT

The authors received no specific funding for this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest to report regarding the present study.

AUTHOR CONTRIBUTIONS

Conceptualization, H.A, A.A, and S.H.; methodology, S.H.; software, H.A, A.A, G.R. and H.S.; validation, S.H., H.S. and G.R.; writing—original draft preparation, H.A., A.A; writing—review and editing, S.H. All authors have read and agreed to the published version of the manuscript.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

In this research Kaggle dataset AUITD is used.

Link: https://www.kaggle.com/datasets/azouzmaroua/algeria-ultrasound-images-thyroid-dataset-auitd

Abbreviations

AI	Artificial Intelligence
AUITD	Algeria Ultrasound Images Thyroid Dataset
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CAD	Computer Aided Diagnosis
DL	Deep Learning
FTC	Follicular Thyroid Carcinoma
KNN	K-nearest Neighbor
ML	Machine Learning
SSD	Solid-State Drive
SVM	Support Vector Machine
VGG16	Visual Geometry Group-16

REFERENCES

[1] E. A. Agyekum *et al.*, "Ultrasound-based classification of follicular thyroid Cancer using deep convolutional neural networks with transfer learning," *Sci. Rep.*, vol. 15, no. 1, p. 21708, 2025.

[2] M. Krönke *et al.*, "Tracked 3D ultrasound and deep neural network-

based thyroid segmentation reduce interobserver variability in thyroid volumetry," *PLoS One*, vol. 17, no. 7, p. e0268550, 2022.
[3] T. Hu *et al.*, "Machine Learning-Based Prediction of Lymph Node Metastasis and Volume Using Preoperative Ultrasound Features in Papillary Thyroid Carcinoma," *J. Clin. Ultrasound*, vol. 54, no. 2, pp. 293–303, 2026.
[4] J. Weng *et al.*, "Deep learning for classification of thyroid nodules on ultrasound: validation on an independent dataset," *Clin. Imaging*, vol. 99, pp. 60–66, 2023.
[5] Y.-J. Kim *et al.*, "Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists," *Eur. J. Radiol.*, vol. 152, p. 110335, 2022.
[6] P.-S. Zhu *et al.*, "Ultrasound-based deep learning using the VGGNet model for the differentiation of benign and malignant thyroid nodules: a meta-analysis," *Front. Oncol.*, vol. 12, p. 944859, 2022.
[7] F. Jerbi, N. Aboudi, and N. Khlifa, "Automatic classification of ultrasound thyroids images using vision transformers and generative adversarial networks," *Sci. African*, vol. 20, p. e01679, 2023.
[8] J. Wang *et al.*, "Thyroid ultrasound diagnosis improvement via multi-view self-supervised learning and two-stage pre-training," *Comput. Biol. Med.*, vol. 171, p. 108087, 2024.
[9] J. Kim *et al.*, "Deep learning technology for classification of thyroid nodules using multi-view ultrasound images: potential benefits and challenges in clinical application," *Endocrinol. Metab.*, vol. 40, no. 2, pp. 216–224, 2025.
[10] O. Abdelrazik, M. Elsayed, N. Wahab, N. Rajpoot, and A. Shephard, "A Deep Learning Framework for Thyroid Nodule Segmentation and Malignancy Classification from Ultrasound Images," *arXiv Prepr. arXiv2511.11937*, 2025.
[11] W. Wu *et al.*, "Prediction of Seronegative Hashimoto's thyroiditis using machine learning models based on ultrasound radiomics: a multicenter study," *BMC Immunol.*, vol. 26, no. 1, p. 27, 2025.
[12] Y. Xu, M. Xu, Z. Geng, J. Liu, and B. Meng, "Thyroid nodule classification in ultrasound imaging using deep transfer learning," *BMC Cancer*, vol. 25, no. 1, p. 544, 2025.
[13] Y. Sharifi, M. D. Ashgzari, S. Shafiei, S. R. Zakavi, and S. Eslami, "Using deep learning for thyroid nodule risk stratification from ultrasound images," *WFUMB Ultrasound Open*, vol. 3, no. 1, p. 100082, 2025.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
[17] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

Hafsa Azam is a BSCS student with research interests in Artificial Intelligence, Machine Learning, and Web Development. She has experience in Python and modern computing technologies and is actively involved in academic and research projects.

Alisha Ashraf is currently a BSCS student with a keen focus on Artificial Intelligence, Machine Learning, and Web Development. She possesses hands-on knowledge of Python and contemporary computing tools and actively participates in academic studies and research-based projects.

Dr. Shahzad Hussain holds a PhD in Computer Science and has over 10 years of experience in teaching and research. His areas of expertise include Artificial Intelligence, Machine Learning, Computer Vision, and Medical Imaging.

Ghazanfar Rehman Innovator and researcher in Smart Grids, AI Surveillance, IoT, and Networks. Based in Edmonton, Canada, leading innovation at Central Protection Services, Central Technologies, and Central Laboratories. Focus: building smart, technology-driven security systems for real-world impact.

Engr. Dr. Hammad Shahab has PhD degree in Computer Engineering with experience in IoT smart farming, automation, cloud-based monitoring, and teaching.