

Article

A Proactive Explainable Artificial Neural Network Model for the Early Diagnosis of Thyroid Cancer

Sumayh S. Aljameel 

Department of Computer Science, College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia; saljameel@iau.edu.sa

Abstract: Early diagnosis of thyroid cancer can reduce mortality, and can decrease the risk of recurrence, side effects, or the need for lengthy surgery. In this study, an explainable artificial neural network (EANN) model was developed to distinguish between malignant and benign nodules and to understand the factors that are predictive of malignancy. The study was conducted using the records of 724 patients who were admitted to Shengjing Hospital of China Medical University. The dataset contained the patients' demographic information, nodule characteristics, blood test findings, and thyroid characteristics. The performance of the model was evaluated using the metrics of accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC). The SMOTEENN combined sampling method was used to correct for a significant imbalance between malignant and benign nodules in the dataset. The proposed model outperformed a baseline study, with an accuracy of 0.99 and an AUC of 0.99. The proposed EANN model can assist health care professionals by enabling them to make effective early cancer diagnoses.

Keywords: artificial neural network; explainable artificial intelligence; thyroid cancer



Citation: Aljameel, S.S. A Proactive Explainable Artificial Neural Network Model for the Early Diagnosis of Thyroid Cancer. *Computation* **2022**, *10*, 183. <https://doi.org/10.3390/computation10100183>

Academic Editors: Demos T. Tsahalidis and Gennady Bocharov

Received: 28 August 2022

Accepted: 7 October 2022

Published: 11 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Thyroid cancer is the 13th most common cancer diagnosis among adolescents and young adults, with 43,800 cases expected in 2022 [1,2]. It accounts for more than 90% of endocrine system malignancies and ranks as the sixth most common among women [1]. Follicular cells are responsible for 96% of thyroid tumors, and 99% of these are differentiated thyroid cancers (DTC) [3]. The main treatments for DTC are surgery, radioactive iodine remnant ablation, and TSH-suppressive therapy with levothyroxine [4]. Each patient's treatment plan is unique and dependent on the lesion's characteristics, but surgery is the most common option [5]. The surgical procedure places an emphasis on identifying benign and malignant thyroid nodules. An accurate prior diagnosis reduces the risk of recurrence, reduces side effects, and allows surgery to proceed more quickly [6]. Thorough post-surgery care also aids in improving survival rates. As a result, it is essential to make precise diagnoses and predictions using thyroid ultrasonography, blood work, and other clinical data [6,7]. Presently, physicians' clinical judgment is used to determine whether a nodule is malignant, but this can be laborious and error prone. To improve treatment decisions and minimize labor, accurate and comprehensive prediction models are urgently required.

Artificial intelligence (AI) is a field of computer science that develops techniques for machines to use knowledge to solve problems. AI is commonly used in the healthcare industry and can contribute to creating new treatments, selecting treatments for complex illnesses, improving the course of care for patients with chronic diseases, and reducing medical mistakes [8]. However, these applications have some shortcomings, such as the "black box" nature of some AI models: because neither the underlying learning mechanisms nor the output models are fully understood, it can be difficult for medical professionals to understand these models and draw clinical conclusions that can be explained to patients and

colleagues. Transparency in medical AI applications is necessary for clinicians and patients to trust them. In recent years, interest in the study of explainable artificial intelligence (EAI) has grown rapidly [9], but to the author's knowledge few studies have investigated the use of EAI for the early detection of thyroid cancer.

In this study, an EAI model is developed, which is capable of diagnosing thyroid nodules with better performance than the benchmark study [10]. Crucially, the model's output can be readily explained.

2. Related Studies

Physicians can benefit from AI in a range of scenarios. Machine learning is widely used in healthcare for the diagnosis of diseases, the development of new drugs, and the determination of patient risk factors. To detect diseases using AI approaches, a variety of medical data sources, including ultrasound images, MRI, and medical records are needed. In recent years, AI techniques have led to significant improvements in healthcare systems and have been employed to accelerate the process of getting patients ready to continue their recovery at home [11].

2.1. AI-Based Studies on the Early Diagnosis of Thyroid Cancer

Chan et al. [12] used a deep learning technique, Deep Convolutional Neural Networks (CNN), to diagnose DTC. The dataset used in this study consisted of ultrasound images of 421 DTC patients and 319 patients with benign tumors. Three CNN methods were used for training the dataset: InceptionV3, ResNet101, and VGG19. The results showed that ResNet101 outperformed the other methods with an accuracy of 77.6%, while InceptionV3 and VGG19 achieved accuracies of 76.5% and 76.1%, respectively. Similarly, Naglah et al. [13] used a computer-aided system to distinguish between benign and malignant thyroid nodules. Their model was developed using the deep learning technique CNN and was evaluated using the magnetic resonance images of 49 patients. The model achieved an accuracy of 87%, a specificity of 97%, and a sensitivity of 69%.

Teknologi et al. [14] proposed a diagnostic for thyroid follicular carcinoma and its subtypes. This study used biopsy images from Pakistani hospitals. An ensemble method was applied using multiclass support-vector machines (SVM) to develop the model. The results showed that the model successfully diagnosed cancers with an accuracy of 98.5%. Olatunji et al. [15] developed a machine learning-based model to detect thyroid cancer by comparing three classifiers: random forest (RF), artificial neural network (ANN), SVM, and naïve Bayes. These classifiers were compared using a dataset collected from Saudi hospitals. The RF model outperformed other models with an accuracy of 90.91% and used only seven of the fifteen measured features of the tumors. In a similar vein, Yang et al. [16] used the machine learning classifiers RF, SVM, and decision tree with genomic data for the early diagnosis of thyroid cancer. The model was evaluated using 506 patients using 703 RNA features. The highest accuracy achieved was 99.2% using RF.

Zhao et al. [17] developed a machine learning-based model using three classifiers: RF, logistic regression, SVM, gradient boost, and ANN. This model was evaluated using 177 ultrasound images with 10 features: shape, composition, size, margins, micro-calcification, halo sign, strain ratio, the echogenicity of the solid portion, vascularity, and the color scale scoring system of real-time elastography. RF outperformed other classifiers with an accuracy of 86% and an area under the curve (AUC) of 93.4%.

2.2. Applications of EAI in Diagnosis and Surgery

Explainability is the extent to which the reasoning behind AI's decisions can be communicated in language that can be understood by a wide range of end users. A concept closely related to explainability is interpretability. Explainability focuses on how the AI's decisions are explained in novel cases, whereas interpretability refers to how the model is understood after being trained on data [18]. There have been several recent studies applying EAI to medicine.

Aslam [18] proposed an EAI approach for the early prediction of ventilator support and mortality in COVID-19 patients. This study used data on patients hospitalized with COVID-19 from King Abdulaziz Medical City in Riyadh. Additionally, the synthetic minority oversampling technique (SMOTE) was used as the dataset suffered from imbalance. The balanced accuracy of the model was 98%, with an AUC of 99.8%, for predicting mortality. For predicting ventilator support, the model achieved a balanced accuracy of 97.9% and an AUC of 98.1%.

El-Sappagh et al. [19] developed an EAI detection and prediction model for Alzheimer's disease. The dataset included 1048 patients, of whom 294 were cognitively normal, 254 had stable mild cognitive impairment, 232 had progressive mild cognitive impairment, and 268 had Alzheimer's disease. The study used RF as a classifier on a two-layer network. The model performed a multi-class classification for the early diagnosis of Alzheimer's patients in the first layer. The model used binary classification in the second layer to look for potential progression from mild impairment to Alzheimer's within three years after a baseline diagnosis. In terms of explainability, Shapley Additive Explanations (SHAP) was used for the RF classifier. The explanations were delivered in plain language so that they were easily understood by doctors. The model achieved 93.95% cross-validation accuracy and an 93.94% F1 score in the first layer, and 87.08% cross-validation accuracy and an 87.09% F1 score in the second layer.

Chen et al. [20] proposed an EAI model using Bayesian networks and CNN for an interpretable clinical diagnosis system. The authors used annotated medical records from multiple hospitals. The results demonstrated that the proposed framework performed more accurately than the prior automatic diagnosis techniques, and its diagnosis explanation was understandable. Magesh et al. [21] developed a machine learning model based on EAI for the diagnosis of Parkinson's disease through dopaminergic imaging techniques such as SPECT and DaTSCAN. The model was trained using CNN and explained using the Local Interpretable Model-Agnostic Explainer (LIME). The results showed that the model successfully detected Parkinson's disease with an accuracy of 95.2%, a sensitivity of 97.5%, and a specificity of 90.9%.

Similarly, Aghamohammadi et al. [22] developed a prediction model for diagnosing heart attacks using EAI techniques. They combined a genetic algorithm with an adaptive fuzzy inference system. Explainable graphs were offered to address the forecasts' explainability. The proposed algorithm performed well, and the study demonstrated that several signs are crucial for the accurate prediction of heart attacks.

3. Materials and Methods

All computer analysis in this study was performed using Python version 3.6 (Python Software Foundation, Wilmington, DE, USA).

3.1. Exploratory Dataset Analysis

The dataset used in this study was the same as that used by Xi et al. [10] and consisted of the records of 724 patients who were admitted to Shengjing Hospital of China Medical University. All participants received a thyroidectomy, and the dataset included the following 19 attributes:

- Patient demographics: gender and age;
- Nodule characteristics: size, site (right, left, or isthmus), multifocality (unifocal or multifocal—i.e., whether there were multiple nodules in one location), shape (regular or irregular), calcification (absent or present), ultrasound echo strength (none, isoechoic, medium-echogenic, hyperechogenic, or hypoechogenic), margin (clear or unclear), blood flow (normal or enriched), composition (cystic, mixed, or solid), laterality (unilateral or multilateral), and malignancy (benign or malignant);
- Blood test findings: free triiodothyronine (FT3), free thyroxine (FT4), thyroid-stimulating hormone (TSH), thyroid peroxidase antibodies (TPO), and thyroglobulin antibodies (TgAb);

- Thyroid characteristics: echo pattern (even or uneven).

Only the largest nodule was included in the dataset if the patient had numerous nodules in a single location—the dataset contained 1232 nodules in total. The target attribute was malignancy, which has two possible values: benign and malignant. Figure 1 shows the number of nodules for each of these outcomes.

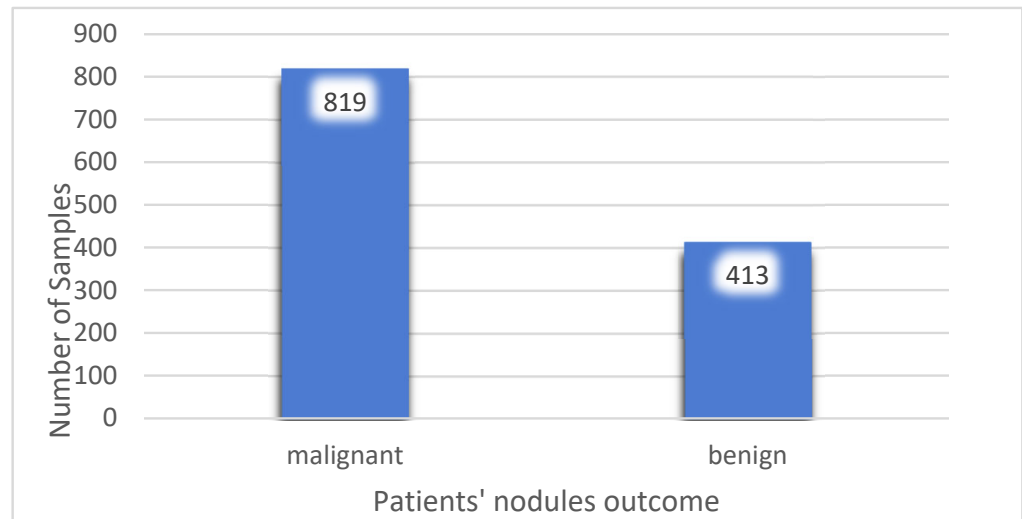


Figure 1. The number of nodules samples per outcome.

The ages of the patients ranged from 13 to 82, and the mean age was 46.6. Figure 2 is a representation of the nodule data binned by patient ages, while in Figure 3 the data are binned according to patient age and nodule malignancy, and in Figure 4 the nodules are also binned by patient data. These plots demonstrate that the modal age group was 43–52 and that there were far more nodules from female patients (1032) than male patients (200). Figure 5 shows the correlation values for all 19 features, and it shows a significant correlation between the nodule outcome and four features: site, shape, calcification, blood flow, and margin.

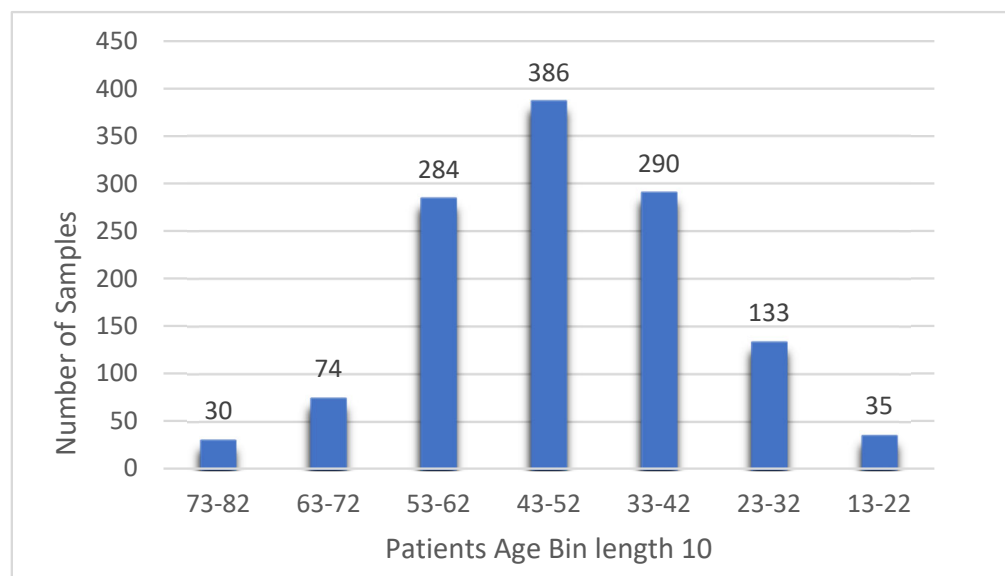


Figure 2. Patient age distribution according to the number of the samples.

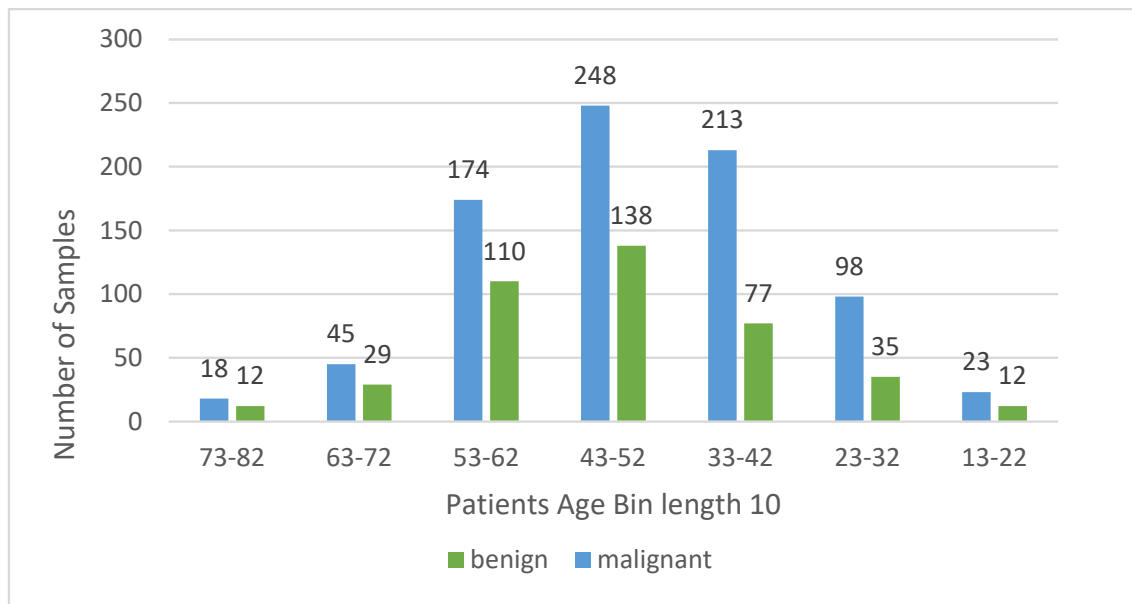


Figure 3. Distribution of patient age range according to patients' nodule outcomes.

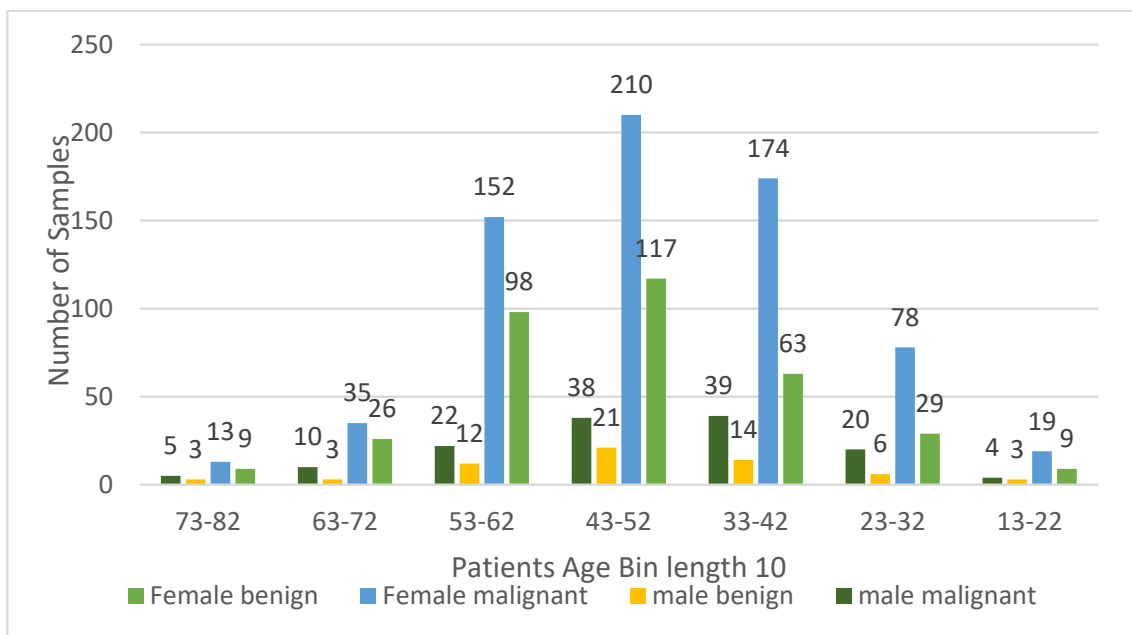


Figure 4. Distribution of patient age range according to gender and patients' nodule outcomes.

3.2. Artificial Neural Network Model

An Artificial Neural Network (ANN) is a simulated network of neurons designed to emulate the human brain [23]. ANNs may learn without supervision, unlike most other AI techniques. Users simply need to provide the ANN with training data—they do not need to give it specific instructions on how to make decisions, so no complex programming is required [24]. Many diseases currently have unknown causes, and the symptoms of some conditions evolve over time. Diagnosis and treatment decisions are typically based on the experience and considered judgment of medical professionals. As a result, an ANN's learning, memory, and induction capabilities define how well-suited it is for use in medicine [25].

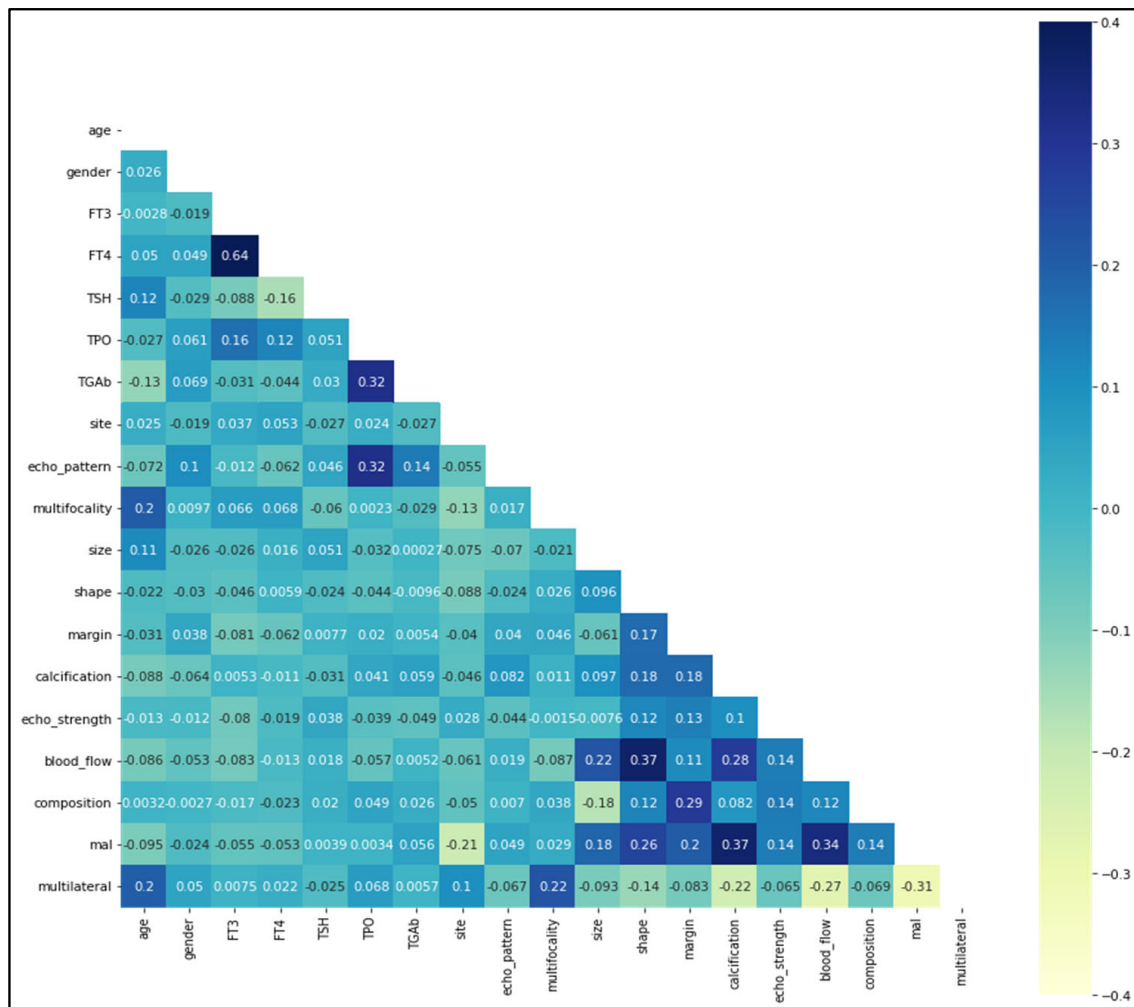


Figure 5. The data correlation of the 19 features.

ANNs have four key advantages: their capacity for self-learning, their ability to conduct a large number of processes concurrently, their strong promotion ability to generalize and predict on unobserved data after learning from the original inputs and their associations, and their capacity to synthesize information [26]. Conceptually, an ANN can be split into three layers: the input layer, in which data enters the network; the hidden layer, which derives patterns from the data; and the output layer, which uses these patterns for the data classification or regression [26].

In this study, two ANN models were developed; the first model includes all dataset features, while the second model is limited to 10 selected features, which are selected based on the highest correlations values. The neural tangent kernel (NTK) was used to implement the models. NTK uses gradient descent to explain how a deep ANN evolved. A sequential method was used to initialize the model and add the layers. In the first model with the full set of features, the input layer has 18 neurons, whereas the second model has an input layer of 10 neurons. Both models have five hidden layers of 512, 256, 128, 64, and 32 neurons, respectively. For the activation function, the rectified linear unit (ReLU) was used, which processes the inputs by converting any negative values to zero:

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

In both models, the output layer has one neuron and uses the sigmoid activation function:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

To avoid overfitting, dropout layers were added after the third layer with a rate of 20%. The Adam optimization algorithm [27] was used for training deep learning models. Binary and categorical cross-entropy were used to calculate the loss, while the accuracy metric was used to evaluate the model. The training process was carried out using optimum values of the hyperparameters of epochs and batch size, which were found using a grid search optimization algorithm [28]. Specifically, 100 epochs and a batch size of 10 were used.

3.3. Explainability of the Proposed Model

An explainable AI was added to the proposed model to study its interpretability. SHAP values [29] were used to quantify the size of the impact of the various features on the model, and for the second model, these are shown in Figure 6.

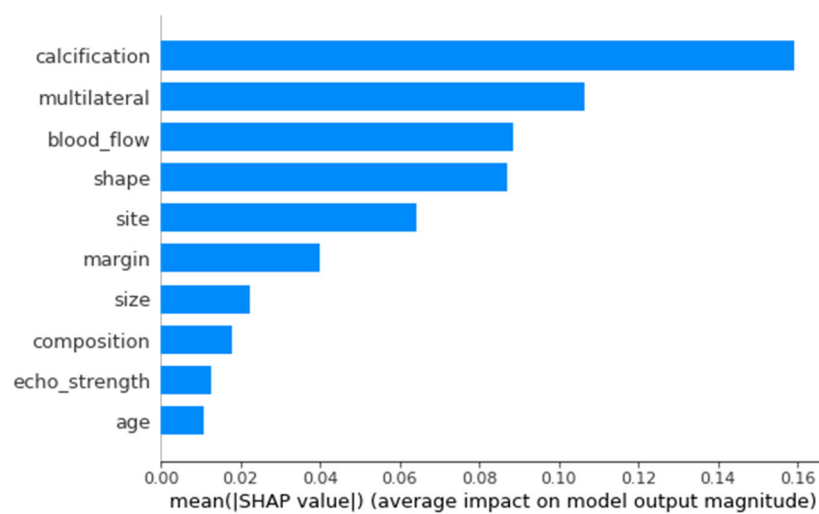


Figure 6. Mean SHAP values for the features of the second model.

A decision tree was created to further enhance the interpretability of the model. Figure 7 shows the decision tree for the classification of benign and malignant thyroid cancers. For each decision level, the node shows the following information:

- The name of the feature used;
- The number of samples used;
- The number of samples at the node that fall into each class;
- The Gini score: a metric that quantifies the purity of the node. A Gini score below zero means that the samples in each leaf belong to a single class.

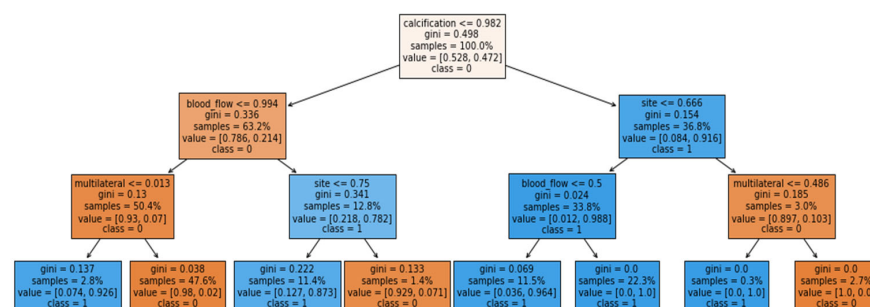


Figure 7. A decision tree representation of the ANN model. Here “class” represents the malignancy of the cancer, with a value of 0 for a benign cancer and 1 for a malignant cancer. “value” represents the fraction of samples that fall into each class.

3.4. Evaluation Measures

The performance of the model was evaluated using the metrics of accuracy, sensitivity, specificity, precision, F1 score, and AUC.

Accuracy is simply the percentage of samples that were classified correctly. In a medical context it can be expressed in terms of the numbers of true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*) as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Sensitivity is a measure of how good the model is at picking up positive cases, and is calculated as the percentage of positive cases that were classified correctly:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

Similarly, specificity is a measure of how often the model correctly classifies negative cases:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

Precision is a measure of the quality of the model's positive classifications. It is calculated as the percentage of positive classifications that were correct:

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

The F1 score is the harmonic mean of the precision and the sensitivity. It takes into account both the false positive and false negative rates, with an F1 score close to 1 implying that both are low.

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (7)$$

The AUC value was also used to measure the performance of the model. Its value varies between 0 and 1, with higher values corresponding to better performance.

4. Results

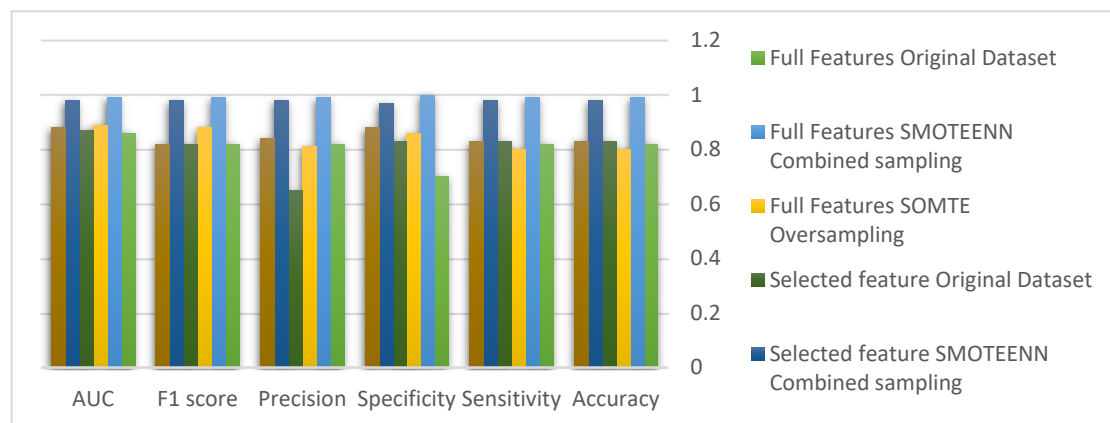
The two different models (one using all features, the other using 10 selected features) were trained on the data, and their performance was compared using several evaluation measures.

The dataset suffered from an imbalance—there were far more malignant nodules than benign ones—so it was necessary to use a technique to avoid biasing the models toward one of the outcomes. The SMOTEENN algorithm [30] was employed for this purpose. This is a combination of oversampling using SMOTE and undersampling using the Edited Nearest Neighbors (ENN) method. The SMOTEENN algorithm starts by applying the SMOTE algorithm first which is a statistical method for evenly expanding the number of instances in the dataset. For each target class and its near neighbors, the algorithm samples the feature space. After that, the algorithm creates new instances by incorporating features from the target case. Then the ENN algorithm is applied by finding each observation's K-nearest neighbor first, then determining whether or not the majority class from the observation's K-nearest neighbor matches the observation's class. The observation and its K-nearest neighbor are removed from the dataset if the majority class of the observation's K-nearest neighbor and the observation's class vary. The dataset was split and evaluated using the K-fold cross-validation technique ($K = 10$) in order to avoid any bias in the training and testing sets. This involved splitting the data into ten subsets, using nine subsets for training, and leaving the last one for testing the trained model. The training subsets were also split into training and validation sets. Table 1 shows the testing results for the proposed model for the diagnosis of thyroid cancer.

Table 1. Results of the proposed model for the diagnosis of thyroid cancer.

Feature Set	Techniques	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
Full Features	Original Dataset	0.82	0.82	0.70	0.82	0.82	0.86
	SMOTEENN Combined sampling	0.99	0.99	1.00	0.99	0.99	0.99
	SOMTE Oversampling	0.80	0.80	0.86	0.81	0.88	0.89
Selected features	Original Dataset	0.83	0.83	0.83	0.65	0.82	0.87
	SMOTEENN Combined sampling	0.98	0.98	0.97	0.98	0.98	0.98
	SOMTE Oversampling	0.83	0.83	0.88	0.84	0.82	0.88

The evaluation measures described in Section 3.4 were calculated for the two models using the original dataset and two different data sampling techniques. The highest measures were achieved using the full set of features and the SMOTEENN combined sampling technique, which yielded a score of at least 0.99 for all of the evaluation measures. Figure 8 shows a comparison between the evaluation measures for the two models and the three approaches to data sampling.

**Figure 8.** Evaluation measures for the full model and the selected features model, using the original dataset and two data sampling techniques.

In the baseline study by Xi et al. [10], the best accuracy of 0.79 and AUC of 0.84 was achieved using Random Forest and the original dataset with full features. The present study's model achieved an accuracy of 0.82 and an AUC of 0.86. However, applying the SMOTEENN sampling technique yielded a significantly higher accuracy and an AUC of 0.99. In the case of SOMTE, the model achieved an accuracy of 0.80 and an AUC of 0.89 using the full set of features.

In a medical setting, it is conventional to focus on the sensitivity and specificity, which are the probability of correctly classifying patients with and without the disease, respectively. The highest sensitivity and specificity were achieved using the full features and the SMOTEENN sampling technique—in this case the sensitivity was 0.99 and the specificity was 1.00. Similarly, in the selected features model, the highest sensitivity and specificity values—0.98 and 0.97, respectively—were achieved using the SMOTEENN sampling technique.

5. Discussion

A deep learning model was used to predict the malignancy of thyroid nodules, using previously published data from medical tests. An EAI model was used to explain the impacts of the various nodule features in the deep learning model. The results suggest that nodule calcification, multifocality, and enriched blood flow have the greatest impact in diagnosing malignant nodules. Previous studies have found that nodule calcification is the strongest indicator for malignancy [31–33]. Other studies have found that multifocality

has a significant impact [34–36]. Debnam et al. [36] raised doubts about whether enriched blood flow is a significant diagnostic factor, whereas the results in the present study found that it did have a significant impact based on the EAI model.

The site of the nodule is also considered an important feature by the proposed model. This aligns with many prior studies that have concluded that the thyroid nodule site is an important indicator of thyroid cancer, with isthmic nodules being particularly likely to be malignant [37–39].

6. Conclusions

In this study, an EANN model was developed to diagnose malignant thyroid nodules. The model used a dataset of 724 patients including their clinical and demographic information—19 features in total, including the target attribute of malignancy, which has two possible values: benign and malignant. The model was trained on this dataset and optimized using the Adam optimizer.

To evaluate the model, six measurements were used: accuracy, sensitivity, specificity, F1 score, precision, and AUC. As the dataset suffered from imbalance, the SMOTEENN balancing technique was used. A variant of the model using only 10 of the features was also studied. The results showed that the proposed model outperformed the baseline study, with an accuracy of 0.99 using the full features and 0.98 using the selected 10 features. EAI was employed to interpret the proposed model and to understand the impact of the different features in predicting malignant thyroid nodules. A decision tree was created to specify the rules behind the model.

The proposed model can assist healthcare professionals in diagnosing malignant thyroid nodules and explaining the reasoning behind their diagnoses. In future, this study could be expanded by applying the model to a larger dataset.

Funding: This research received no external funding.

Data Availability Statement: The study used an open-source dataset accessible at https://zenodo.org/record/6387087#.Yj_KeeqZM4c (accessed on 7 July 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [CrossRef] [PubMed]
2. American Cancer Society. Cancer Statistics Center. Available online: <https://cancerstatisticscenter.cancer.org> (accessed on 1 June 2022).
3. Fagin, J.A.; Wells, S.A. Biologic and Clinical Perspectives on Thyroid Cancer. *N. Engl. J. Med.* **2016**, *375*, 1054–1067. [CrossRef] [PubMed]
4. Schmidbauer, B.; Menhart, K.; Hellwig, D.; Grosse, J. Differentiated Thyroid Cancer—Treatment: State of the Art. *Int. J. Mol. Sci.* **2017**, *18*, 1292. [CrossRef] [PubMed]
5. Yeh, M.W.; Bauer, A.J.; Bernet, V.A.; Ferris, R.L.; Loevner, L.A.; Mandel, S.J.; Orloff, L.A.; Randolph, G.W.; Steward, D.L. American Thyroid Association Statement on Preoperative Imaging for Thyroid Cancer Surgery. *Thyroid* **2015**, *25*, 3–14. [CrossRef]
6. Durante, C.; Grani, G.; Lamartina, L.; Filetti, S.; Mandel, S.J.; Cooper, D.S. The Diagnosis and Management of Thyroid Nodules. *JAMA* **2018**, *319*, 914–924. [CrossRef]
7. Avram, A.M.; Zukotynski, K.; Nadel, H.R.; Giovanella, L. Management of Differentiated Thyroid Cancer: The Standard of Care. *J. Nucl. Med.* **2022**, *63*, 189–195. [CrossRef]
8. Shen, J.; Zhang, C.J.P.; Jiang, B.; Chen, J.; Song, J.; Liu, Z.; He, Z.; Wong, S.Y.; Fang, P.-H.; Ming, W.-K. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. *JMIR Med. Inf.* **2019**, *7*, e10010. [CrossRef]
9. Meske, C.; Bunde, E.; Schneider, J.; Gersch, M. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Inf. Syst. Manag.* **2022**, *39*, 53–63. [CrossRef]
10. Xi, N.M.; Wang, L.; Yang, C. Improving The Diagnosis of Thyroid Cancer by Machine Learning and Clinical Data. *arXiv* **2022**, arXiv:2203.15804. [CrossRef]
11. Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F. Artificial Intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework and Future Research Agenda. *J. Ambient. Intell. Hum. Comput.* **2022**, *9*, 1–28. [CrossRef]
12. Chan, W.K.; Sun, J.H.; Liou, M.J.; Li, Y.R.; Chou, W.Y.; Liu, F.H.; Chen, S.T.; Peng, S.J. Using Deep Convolutional Neural Networks for Enhanced Ultrasonographic Image Diagnosis of Differentiated Thyroid Cancer. *Biomedicines* **2021**, *9*, 1771. [CrossRef] [PubMed]

13. Naglah, A.; Khalifa, F.; Khaled, R.; Razeq, A.A.K.A.; Ghazal, M.; Giridharan, G.; El-Baz, A. Novel Mri-Based Cad System for Early Detection of Thyroid Cancer Using Multi-Input CNN. *Sensors* **2021**, *21*, 3878. [\[CrossRef\]](#)
14. Teknologi, J.; Ahmed, J.; Rehman, M.A. Cancer Prevention Initiative: An Intelligent Approach for Thyroid Cancer Type Diagnostics. *J. Teknol.* **2016**, *78*. [\[CrossRef\]](#)
15. Olatunji, S.O.; Alotaibi, S.; Almutairi, E.; Alrabae, Z.; Almajid, Y.; Altabee, R.; Altassan, M.; Basheer Ahmed, M.I.; Farooqui, M.; Alhiyafi, J. Early Diagnosis of Thyroid Cancer Diseases Using Computational Intelligence Techniques: A Case Study of a Saudi Arabian Dataset. *Comput. Biol. Med.* **2021**, *131*, 104267. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Yang, F.; Zhang, J.; Li, B.; Zhao, Z.; Liu, Y.; Zhao, Z.; Jing, S.; Wang, G. Identification of Potential LncRNAs and MiRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning. *Int. J. Endocrinol.* **2021**, *2021*, 3984463. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Zhao, H.-N.; Liu, J.-Y.; Lin, Q.-Z.; He, Y.-S.; Luo, H.-H.; Peng, Y.-L.; Ma, B.-Y. Partially Cystic Thyroid Cancer on Conventional and Elastographic Ultrasound: A Retrospective Study and a Machine Learning—Assisted System. *Ann. Transl. Med.* **2020**, *8*, 495. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zhang, Y.; Weng, Y.; Lund, J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics* **2022**, *12*, 237. [\[CrossRef\]](#)
19. El-Sappagh, S.; Alonso, J.M.; Islam, S.M.R.; Sultan, A.M.; Kwak, K.S. A Multilayer Multimodal Detection and Prediction Model Based on Explainable Artificial Intelligence for Alzheimer’s Disease. *Sci. Rep.* **2021**, *11*, 2660. [\[CrossRef\]](#)
20. Chen, J.; Dai, X.; Yuan, Q.; Lu, C.; Huang, H. Towards Interpretable Clinical Diagnosis with Bayesian Network Ensembles Stacked on Entity-Aware CNNs. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.
21. Magesh, P.R.; Myloth, R.D.; Tom, R.J. An Explainable Machine Learning Model for Early Detection of Parkinson’s Disease Using LIME on DaTSCAN Imagery. *Comput. Biol. Med.* **2020**, *126*, 104041. [\[CrossRef\]](#)
22. Aghamohammadi, M.; Madan, M.; Hong, J.K.; Watson, I. Predicting Heart Attack Through Explainable Artificial Intelligence. In *International Conference on Computational Science*; Springer: Cham, Switzerland, 2019; pp. 633–645.
23. Bui, D.-K.; Nguyen, T.N.; Ngo, T.D.; Nguyen-Xuan, H. An Artificial Neural Network (ANN) Expert System Enhanced with the Electromagnetism-Based Firefly Algorithm (EFA) for Predicting the Energy Consumption in Buildings. *Energy* **2020**, *190*, 116370. [\[CrossRef\]](#)
24. Toghraie, D.; Sina, N.; Jolfaei, N.A.; Hajian, M.; Afrand, M. Designing an Artificial Neural Network (ANN) to Predict the Viscosity of Silver/Ethylene Glycol Nanofluid at Different Temperatures and Volume Fraction of Nanoparticles. *Phys. A Stat. Mech. Its Appl.* **2019**, *534*, 122142. [\[CrossRef\]](#)
25. Berus, L.; Klancnik, S.; Brezocnik, M.; Ficko, M. Classifying Parkinson’s Disease Based on Acoustic Measures Using Artificial Neural Networks. *Sensors* **2019**, *19*, 16. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Wu, Y.; Feng, J. Development and Application of Artificial Neural Network. *Wirel. Pers. Commun.* **2018**, *102*, 1645–1656. [\[CrossRef\]](#)
27. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
28. Khalid, R.; Javaid, N. A Survey on Hyperparameters Optimization Algorithms of Forecasting Models in Smart Grid. *Sustain. Cities Soc.* **2020**, *61*, 102275. [\[CrossRef\]](#)
29. Giudici, P.; Raffinetti, E. Shapley-Lorenz EXplainable Artificial Intelligence. *Expert Syst. Appl.* **2021**, *167*, 114104. [\[CrossRef\]](#)
30. Batista, G.E.A.P.A.; Prati, R.C.; Monard, M.C. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [\[CrossRef\]](#)
31. Peduk, S. Predictive Importance of Ultrasonography and Anti-Thyroid Antibodies in the Management of Thyroid Nodules in Indeterminate Cytology. *Eurasian J. Med. Investig.* **2022**, *6*, 122–127. [\[CrossRef\]](#)
32. Yin, L.; Zhang, W.; Bai, W.; He, W. Relationship Between Morphologic Characteristics of Ultrasonic Calcification in Thyroid Nodules and Thyroid Carcinoma. *Ultrasound Med. Biol.* **2020**, *46*, 20–25. [\[CrossRef\]](#)
33. Kim, H.L.; Ha, E.J.; Han, M. Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography. *Ultrasound Med. Biol.* **2019**, *45*, 2672–2678. [\[CrossRef\]](#)
34. Mazeh, H.; Samet, Y.; Hochstein, D.; Mizrahi, I.; Ariel, I.; Eid, A.; Freund, H.R. Multifocality in Well-Differentiated Thyroid Carcinomas Calls for Total Thyroidectomy. *Am. J. Surg.* **2011**, *201*, 770–775. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Feng, J.-W.; Qu, Z.; Qin, A.-C.; Pan, H.; Ye, J.; Jiang, Y. Significance of Multifocality in Papillary Thyroid Carcinoma. *Eur. J. Surg. Oncol.* **2020**, *46*, 1820–1828. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Debnam, J.M.; Vu, T.; Sun, J.; Wei, W.; Krishnamurthy, S.; Zafereo, M.E.; Weitzman, S.P.; Garg, N.; Ahmed, S. Vascular Flow on Doppler Sonography May Not Be a Valid Characteristic to Distinguish Colloid Nodules from Papillary Thyroid Carcinoma Even When Accounting for Nodular Size. *Gland Surg.* **2019**, *8*, 461–468. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Ramundo, V.; Lamartina, L.; Falcone, R.; Ciotti, L.; Lomonaco, C.; Biffoni, M.; Giacomelli, L.; Maranghi, M.; Durante, C.; Grani, G. Is Thyroid Nodule Location Associated with Malignancy Risk? *Ultrasonography* **2019**, *38*, 231–235. [\[CrossRef\]](#)
38. Jasim, S.; Baranski, T.J.; Teefey, S.A.; Middleton, W.D. Investigating the Effect of Thyroid Nodule Location on the Risk of Thyroid Cancer. *Thyroid* **2020**, *30*, 401–407. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Ma, X.; Xi, B.; Zhang, Y.; Zhu, L.; Sui, X.; Tian, G.; Yang, J. A Machine Learning-Based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr. Bioinform.* **2020**, *15*, 349–358. [\[CrossRef\]](#)