|  |  |
| --- | --- |
| **Title of the paper** | Chaganti, R., Rustam, F., De La Torre Díez, I., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). Thyroid disease prediction using selective features and machine learning techniques. *Cancers*, *14*(16), 3914. |
| **Area of work** | Detection and classification of thyroid disease |
| **Dataset** | Dataset was taken from UCI repository. The dataset contains 9172 sample observations and each sample is represented by 31 features. |
| **Methodology / Strategy** | The dataset consists of several thyroid-related disease records and many target classes. It is followed by the feature selection process, where many feature selection techniques are applied. Experiments are performed with an 80–20 train–test split using several machine learning and deep learning models. |
| **Algorithm** | RF, LR, SVM, ADA, GBM |
| **Result/Accuracy** | Results indicate that extra tree classifier-based selected features tend to provide the highest accuracy of 0.99 when used with the RF model. The lower computational complexity of the machine learning models like RF makes them good candidates for thyroid disease prediction. |

**THYROID DETECTION USING MACHINE LEARNING**

**SUMMARY**

**Paper 1**

**Paper 2**

|  |  |
| --- | --- |
| **Title of the paper** | Gupta, Punit, et al. "Detecting thyroid disease using optimized machine learning model based on differential evolution." *International Journal of Computational Intelligence Systems* 17.1 (2024): 3. |
| **Area of work** | Machine learning approach for thyroid disease detection |
| **Dataset** | The datasets used in this study are taken from the Kaggle repository. The thyroid disease dataset comprises 9172 samples and every sample has 31 features. |
| **Methodology / Strategy** | The dataset contains 25 target classes, of which the top 10 target classes are selected for experiments. The selected targeted dataset is imbalanced, so to make the dataset balanced we used the CTGAN augmentation technique. train machine learning models with a training set and perform hyperparameter optimization using DE optimizer which helps to select the best hyperparameter setting for models. In the end, we evaluate models in terms of accuracy, precision, recall, F1 score, and confusion matrix. |
| **Algorithm** | RF, SVM, LR, AdaBoost, GBM |
| **Result/Accuracy** | RF – 0.98  GBM – 0.92  AdaBoost – 0.99  LR – 0.61  SVM – 0.61 |

**Paper 3**

|  |  |
| --- | --- |
| **Title of the paper** | Hossain, M. B., Shama, A., Adhikary, A., Raha, A. D., Uddin, K. A., Hossain, M. A., ... & Bairagi, A. K. (2023). An explainable artificial intelligence framework for the predictive analysis of hypo and hyper thyroidism using machine learning algorithms. *Human-Centric Intelligent Systems*, *3*(3), 211-231. |
| **Area of work** | compare different classification algorithms used in machine learning |
| **Dataset** | The data was taken from the UCI machine learning repository. Dataset contains 3221 instances with a total of 30 features. |
| **Methodology / Strategy** | Compare and study different classification algorithms used in machine learning |
| **Algorithm** | Decision Tree Classifer, Random Forest Classifer, Naive Bayes Classifer, Gradient Boosting Classifer, Logistic Regression Classifer, K- Nearest Neighbor, Support Vector Machine |
| **Result/Accuracy** | Random Forest (RF) is giving the maximum evaluation score in all sectors in our dataset, and Naive Bayes is performing very poorly. Moreover selecting the feature by using the feature importance method RF provides the best accuracy of 91.42%, precision of 92%, recall of 92% and F1-score of 92%. |

**CONCLUSION**

From the above three papers, we get to know that different approaches are used for thyroid detection and classification. From the observation it is found out that Random Forest is more accurate than the other algorithms.