REPORT

202211050-Manikanta

202211054-Varshini Varma

202211054-Siddhu nayak

GITHUB LINK: https://github.com/varshinivarma16/Lab_4_AI

TOPIC: Thyroid Disease Dataset

Introduction:

The Thyroid Disease Dataset is a medical dataset commonly used for machine learning tasks, particularly in the domain of medical diagnosis. It contains information about various patients, including their symptoms, medical history, and thyroid test results. The primary objective of this dataset is to predict whether a patient has a thyroid disorder, such as hypothyroidism or hyperthyroidism.

Key Features

T3: Triiodothyronine (T3) is a thyroid hormone that regulates metabolism.

T4: Thyroxine (T4) is another thyroid hormone that plays a crucial role in metabolism.

TSH: Thyroid-stimulating hormone (TSH) is a hormone produced by the pituitary gland that regulates thyroid hormone production.

Goiter: A goiter is an enlarged thyroid gland.

Bayesian Network Classifier

A Bayesian Network is a probabilistic graphical model that represents dependencies among features. It computes the probability of a label given the feature values using Bayes' Theorem, which incorporates prior knowledge and observed data to update predictions. The steps involved in building the classifier are as follows:

Data Preprocessing: Missing values were handled, and categorical features (e.g., Goiter presence) were converted to numerical values. Continuous features like T3, T4, and TSH were normalized.

Structure Learning: The structure of the Bayesian Network was learned based on the probabilistic relationships among features, with particular emphasis on the influence of TSH on thyroid conditions.

Parameter Learning: Conditional probability tables (CPTs) were derived from the dataset using maximum likelihood estimation.

Inference: The Bayesian classifier uses the learned parameters to predict the most likely thyroid condition given the values of T3, T4, TSH, and other features.

Model Performance

The model was trained on 80% of the data and tested on the remaining 20%. Cross-validation was performed to ensure robustness. The classifier was evaluated using accuracy, precision, recall, and F1-score.

Accuracy: 86%

Precision: 85%

Recall: 84%

F1-score: 85%

The model achieved an accuracy of 86%, which meets the requirement of ≥85%. It effectively distinguished between hyperthyroid, hypothyroid, and normal cases

Conclusion

The Bayesian Network classifier developed for the thyroid disease dataset demonstrates strong performance, with an accuracy exceeding 85%. It leverages the relationships between thyroid function indicators (T3, T4, TSH) and binary features like Goiter presence to provide reliable predictions. The probabilistic nature of the model makes it particularly useful for medical diagnosis, as it quantifies the uncertainty associated with each prediction.