Variable Comparison Report

Thyroid Dataset vs. Physician-Required Variables for CDSS

Clinical Decision Support System — Hyperthyroidism Management

3	23	18	5
Kaggle Datasets Merged	Variables Available	Variables Missing	Variable Categories

1. Variables Available in the Current Dataset

The unified dataset, merged from three Kaggle thyroid datasets, contains the following variables organised across five categories:

Category	Variables
A. Demographic	Age, Sex
B. Laboratory (Thyroid Function)	TSH, T3, TT4, T4, T4U, FTI (Free Thyroxine Index), TSH measured, T3 measured, TT4 measured, T4U measured
C. Clinical Indicators	Goitre, Nodule size, Family history (general), Diabetes, Obesity
D. Treatment-Related	On antithyroid medication, I131 treatment (radioiodine therapy)
E. Classification / Target	Query hyperthyroid, Query hypothyroid, Query on thyroxine, Class (categorical diagnostic label)

2. Variables Required by the Physician

The supervising physician requested a structured set of variables for developing a comprehensive Clinical Decision Support System (CDSS) for hyperthyroidism management, covering five domains:

Domain	Required Variables
A. Anamnesis	Age, Sex, Smoking status, Weight, Height, BMI, Personal history of autoimmune diseases, Family history of Graves' disease, Family history of Hashimoto's thyroiditis, Detailed autoimmune background

B. Clinical Severity	Frank hyperthyroidism, Moderate hyperthyroidism, Asymptomatic (subclinical) hyperthyroidism
C. Paraclinical Investigations	TSH, FT3, FT4, TRAb/TRAK antibodies, Anti-TPO, Anti-Tg, Thyroid volume, Ultrasound characteristics, Doppler vascular velocity, Scintigraphy pattern
D. Treatment Details	Antithyroid drug type (Methimazole/Carbimazole), Treatment protocol (titration or block-and-replace), Treatment duration, Treatment-induced hypothyroidism
E. Follow-Up & Outcomes (12 months)	Remission, Relapse, Time to relapse

3. Variables Missing from the Current Dataset

Although key variables (Age, Sex, TSH) are present, a significant number of physician-requested variables are absent. The gaps are detailed below by domain:

A. Missing Anamnesis Variables

- ✗ Smoking status
- ✗ Weight and Height
- ✗ Exact BMI calculation
- ✗ Detailed personal autoimmune history
- ✗ Disease-specific family history (Graves' disease, Hashimoto's thyroiditis)

B. Missing Clinical Severity Variable

- ✗ Explicit severity classification (Frank / Moderate / Asymptomatic subclinical)

C. Missing Paraclinical Variables

- ✗ FT3 (Free Triiodothyronine)
- ✗ FT4 (Free Thyroxine)
- ✗ TRAb / TRAK antibodies
- ✗ Anti-TPO antibodies
- ✗ Anti-Tg antibodies
- ✗ Thyroid volume (exact measurement)
- ✗ Thyroid ultrasound characteristics
- ✗ Doppler vascular velocity
- ✗ Scintigraphy pattern / results

D. Missing Treatment Details

✗ Antithyroid drug type
✗ Treatment protocol specification
✗ Treatment duration
✗ Occurrence of treatment-induced hypothyroidism

E. Missing Follow-Up Data
✗ Remission status
✗ Relapse occurrence
✗ Time to relapse

4. Full Side-by-Side Comparison Matrix

Domain	Variable	In Dataset	Required by MD	Status
Demographic	Age	✓	✓	**Available**
	Sex	✓	✓	**Available**
	Weight / Height / BMI	✗	✓	**Missing**
	Smoking status	✗	✓	**Missing**
Laboratory	TSH	✓	✓	**Available**
	T3 / TT4 / T4 / T4U / FTI	✓	Partial	**Partial**
	FT3 (Free T3)	✗	✓	**Missing**
	FT4 (Free T4)	✗	✓	**Missing**
Immunological	TRAb / TRAK antibodies	✗	✓	**Missing**
	Anti-TPO antibodies	✗	✓	**Missing**
	Anti-Tg antibodies	✗	✓	**Missing**
Imaging	Thyroid volume	✗	✓	**Missing**
	Nodule presence	✓	✓	**Available**
	Ultrasound / Doppler / Scintigraphy	✗	✓	**Missing**
Treatment	Antithyroid medication (binary)	✓	Partial	**Partial**
	I131 radioiodine therapy	✓	✓	**Available**
	Drug type / Protocol / Duration	✗	✓	**Missing**
Follow-Up	Remission / Relapse / Time to relapse	✗	✓	**Missing**
	Severity classification	✗	✓	**Missing**

5. Summary & Recommendations

The current unified thyroid dataset provides essential demographic and hormonal variables sufficient for basic hyperthyroidism detection and classification tasks. Key available variables include Age, Sex, TSH, and basic thyroid hormone measurements (T3, TT4, FTI), along with binary clinical flags for treatment history and nodule presence. However, the dataset exhibits significant gaps when evaluated against the physician's requirements for a comprehensive Clinical Decision Support System (CDSS). Critical missing elements include free hormone fractions (FT3, FT4), immunological markers (TRAb, Anti-TPO, Anti-Tg), detailed imaging parameters (thyroid volume, Doppler velocity, scintigraphy), and all follow-up outcome data (remission, relapse, time to relapse). Treatment information is limited to binary flags without drug type, protocol, or duration. To build a clinically meaningful CDSS capable of supporting treatment decisions and long-term prognosis, data collection should be extended through prospective clinical data acquisition or linkage with hospital electronic health records.

● Available ● Partial ● Missing