Medical Expert Validation Form

For the Synthetic Dataset: "Lung Cancer Bangladesh"

Researcher Information

- **Name:** Nishat Vasker
- **Email:** nishatvasker@gmail.com
- **Date:** 4/22/2025

Section 1: Medical Expert Information

- **Name:** DR. NAYAN MANDAL
- **Designation:** Indoore Medical Officer
- **Affiliation:** Sir Salimullah Medical college & Mitford Hospital
- **Email:** nayanmandalssmc42@gmail.com
- **Years of Experience in Lung Cancer Diagnosis:** 1 year

Section 2: Dataset Validation

A. Dataset Overview

This dataset, titled **"Lung Cancer Bangladesh,"** is a **fully synthetic** dataset generated based on statistical patterns observed in real-world lung cancer cases.

- It includes demographic, clinical, environmental, and lifestyle features relevant to lung cancer diagnosis.
- The dataset is designed to train AI models for lung cancer prediction and severity assessment.
- No real patient data has been used, ensuring privacy and ethical compliance. However, the data is **related to real-life medical patterns and distributions** to maintain clinical relevance.

B. Feature Relevance

1. Are the selected features (e.g., age, smoking status, tumor size, cancer stage) clinically relevant for lung cancer diagnosis?
   ☑ Yes ☐ No (If no, please specify missing or irrelevant features)
   Comments: _____
2. Are the class labels (e.g., cancer stages, severity levels) correctly defined in a clinically meaningful way?
   ☑ Yes ☐ No (If no, please specify)
   Comments: _____

C. Dataset Structure & Realism

3. Does the dataset structure (e.g., patient attributes, feature distributions) align with real-world lung cancer cases?

☑ Yes ☐ No

Comments: _____

4. Do the generated patient cases resemble realistic clinical scenarios?

☑ Yes ☐ No

Comments: _____

D. AI Model Interpretability & Trustworthiness

5. Do you believe that integrating Explainable AI (e.g., SHAP, LIME) improves trust in AI predictions for lung cancer?

☑ Yes ☐ No

Comments: _____

6. Would you consider using an AI-based prediction tool trained on this dataset as a **preliminary screening** aid (not a replacement for diagnosis)?

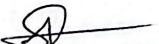☑ Yes ☐ No

Comments: _____

E. Additional Recommendations

7. What changes or improvements would you suggest to make this dataset more medically accurate and relevant?

Comments: _____

Section 3: Expert Declaration & Signature

I, [Expert's Name] _DR. NAYAN MANDAL_ have reviewed the synthetic dataset and provided feedback based on my clinical expertise.

Signature: _____

Date: ___6/9/2025___

Dr. Nayan Mandal
MBBS (SSMC), BCS (Health)
Indoor Medical Officer
Department of Medicine
Sir Salimullah Medical College
& Mitford Hospital, Dhaka