**ADITI GUPTA**

(E23CSEU0401)

PROBLEM STATEMENT

Cancer is becoming a major cause of death as the population grows. It can start anywhere in the body because the body is made of trillions of cells. Normally, cells grow, divide, and die in a controlled way, with new cells replacing old or damaged ones. But in cancer, this process goes wrong. Cancer cells don't die as they should, and new cells grow when they aren't needed. This leads to the formation of extra cells, which can form a mass called a tumor.

Lung cancer, for example, is one of the most common types, and its occurrence has been rising since the 19th century. The causes of lung cancer include smoking, being exposed to radon gas, breathing in secondhand smoke, and exposure to asbestos.

It is one of the deadliest forms of cancer worldwide, and early detection can make a big difference in treatment and survival. However, spotting lung cancer early isn’t always easy, as symptoms might not show up until the cancer has progressed. In this project, I use machine learning to help predict lung cancer risk based on patient data, aiming to support doctors in identifying high-risk patients sooner.

Our goal is to build a computer model that can analyze information about a patient—like their age, lifestyle habits, or other health details—and predict if they might be at risk for lung cancer. This type of model can help doctors focus on patients who might need extra screening or testing, potentially catching the cancer earlier when treatment is more effective.

This project is unique because it goes beyond simple predictions by carefully selecting the most relevant information for lung cancer and testing different machine learning approaches to find the best fit. I also focus on minimizing errors, as both false positives (indicating cancer when there isn’t any) and false negatives (missing an actual case) can have serious consequences. By tuning our model this way, we aim to make it as reliable as possible for real-world use.



1. **Extra Care with Predictions**:   
   My project tries hard to make safe and accurate predictions. It’s designed to avoid giving “false alarms” (saying someone has cancer when they don’t) and to catch real cases when they happen. This care is important in healthcare, where every prediction matters.
2. **Only the Most Important Information**:   
   Instead of using all the data available, my project picks out the most helpful information for predicting lung cancer. By using only the key details, the model can make faster and more accurate predictions, without any extra or confusing data.
3. **More Than Just Accuracy**:   
   Many projects just look at accuracy, but my also checks how well it can find real cases and avoid mistakes. These extra checks make my model more trustworthy for use in healthcare.
4. **Trying Different Methods**:   
   My project doesn’t just rely on one way to make predictions. It tests different methods to find which one works best. This way, I know you’re using the approach that’s most accurate for lung cancer prediction.
5. **Made for Real-Life Use**:   
   My project is built to be helpful in real life. It’s designed so doctors could actually use it to spot patients who might need more testing for lung cancer. This focus on real-life use makes it valuable, not just a project for testing.

