## **Presentation Story Arc & Script Guide**

### **🎬 SLIDE 1: Title Slide (15 seconds)**

**Opening Hook:**

"Good morning everyone. Imagine you're a radiation oncologist treating a head and neck cancer patient. You've just completed their treatment—months of radiation, chemotherapy, countless side effects. And then, a year later, the cancer comes back. **This happens in 50-60% of our patients.** What if we could predict this *before* we even started treatment? That's exactly what we set out to do."

**Why this works:** Starts with a relatable clinical scenario, uses a powerful statistic, and poses an intriguing question.

### **🎬 SLIDE 2: Background & Problem (30 seconds)**

**Story:**

"Locoregional recurrence remains one of the biggest challenges in head and neck cancer. Despite our best treatments, **half of our patients will have the cancer return**. And here's the problem: our current staging systems—while useful—give us population-level statistics, not individual predictions.

As researchers, we wanted to find a better way. But we faced a classic challenge: **we have 103 imaging features from CT scans** and only **163 patients**—that's what we call a 'p much greater than n' problem. Too many variables, too few patients. This leads to unstable models that don't work on new patients."

### **🎬 SLIDE 3: Research Question (20 seconds)**

**Transition:** "So we asked ourselves..."

**Story:**

"**Can we identify a small, stable set of features**—things we can actually understand and trust—that predicts who will have recurrence?

We tested **3 different feature selection methods** against **5 different machine learning classifiers**. But here's what made our approach different: we weren't just chasing the highest accuracy. We wanted **stability, generalization, and most importantly—interpretability**. Because if a doctor can't understand *why* the model makes a prediction, they won't trust it."

### **🎬 SLIDE 4: Methods (40 seconds)**

**Transition:** "Here's how we did it..."

**Story:**

"We analyzed **163 head and neck cancer patients** treated right here at CMC Vellore from 2020 to 2024. [Show flowchart]

From their baseline CT scans, we extracted **103 radiomic features**—these capture tumor shape, size, and texture—things invisible to the human eye. We combined these with **8 clinical variables** like age and stage—things doctors already use.

Now here's the critical part: we split our data **80-20**. The training set for building models, the test set locked away—untouched—for final evaluation. **No cheating, no peeking.** This ensures our results are honest.

We tested everything from simple logistic regression to complex random forests, using different feature selection strategies. The question was: which combination works best?"

### **🎬 SLIDE 5: Results (45 seconds)**

**Transition:** "And here's what we found..."

**Story:**

"The winner? A **simple logistic regression** using just **10 features** selected by an optimization algorithm called Grey Wolf Optimizer. [Point to the trophy]

Test AUC: **0.81** [pause for emphasis]. That's strong discrimination. But here's what's even more important: look at the training AUC—**0.79**. [Point to both numbers] They're almost identical. **No overfitting.** This model will actually work on new patients.

Now look at this comparison [show bar chart]. Random Forest? High training performance but drops on test data—classic overfitting. SVM? Terrible with radiomics alone but dramatically improved when we added clinical variables.

**Key finding:** Clinical variables aren't just helpful—they're essential for generalization."

### **🎬 SLIDE 6: The Signature (25 seconds)**

**Transition:** "So what are these 10 magic features?"

**Story:**

"Four clinical variables—age, stage, T stage, location. Things oncologists already think about.

Six radiomics features—two capture tumor **shape**: size and irregularity. Four capture tumor **texture**: heterogeneity, complexity, patterns you can't see with your eyes but the computer can measure.

This isn't rocket science—it's **biological sense combined with mathematical rigor**."

### **🎬 SLIDE 7: Clinical Explainability (35 seconds)**

**Transition:** "But why do these features matter? Let me show you..."

**Story:**

"Take **Age**: [point to card] Older patients have weaker immune systems, more comorbidities—they can't tolerate aggressive treatment as well.

**Stage and T Stage**: More advanced disease means more tumor burden, hypoxic regions that resist radiation.

**Location**: A hypopharynx tumor is harder to access surgically than an oral cavity tumor.

These aren't arbitrary computer picks—**these are biological realities** that every oncologist already considers intuitively. We've just quantified them."

### **🎬 SLIDE 11: Limitations & Future (30 seconds)**

**Transition:** "Now let me be honest about limitations..."

**Story:**

"This is a **single-center study**—CMC Vellore. Does it work in Mumbai? Delhi? Rural India? **We don't know yet.** That's why external validation is critical.

[Point to roadmap] But here's our plan: **Multi-center validation** is already starting. Then decision curve analysis to prove clinical benefit. Eventually, a **randomized trial** where we actually test whether model-guided treatment improves outcomes.

[Point to timeline] We're here [Phase 1]. The journey to clinical practice is clear."

**Key phrase:** "We're being scientifically honest about what we know—and what we still need to prove."

### **🎬 CLOSING (15 seconds)**

**Powerful Finish:**

"So, back to my opening question: **Before we treat, can we tell?** With this 10-feature signature—**yes, we can**. With 81% accuracy. Using interpretable biology. And ready for clinical testing.

The future of head and neck cancer treatment isn't just about better drugs or radiation—it's about **knowing which patient needs what treatment**. **Personalized medicine, driven by data, grounded in biology.**

Thank you. I'm happy to take questions."