Certainly! Here are **viva questions** phrased as a teacher might ask, with clear, student-style answers based on all the information above:

---

1. Teacher: Why did you choose the Wisconsin Breast Cancer Dataset for your project?

Student:
I chose the Wisconsin Breast Cancer Dataset because it is widely used and trusted by researchers for testing machine learning models in medical diagnosis. The data comes from real patients and includes important measurements that doctors actually use. It also has a balanced mix of cancer and non-cancer cases, which helps in building and testing accurate models. Using this dataset also means my results can be compared easily with other studies.

---

2. Teacher: Why did you select only 10 features instead of using all available features in the dataset?

Student:
I selected the 10 mean features because they are the most basic and important measurements that describe a breast tumor. Research shows these features are the most useful for telling the difference between cancer and non-cancer cases. Using only these keeps the model simpler, helps prevent overfitting, and makes it easier for doctors to use the dashboard, without losing prediction accuracy.

---

3. Teacher: Can you explain what SVM is and why you used it?

Student:
SVM, or Support Vector Machine, is a supervised learning algorithm used for classification. It works by finding the best boundary to separate two classes—in this case, benign and malignant tumors. I used SVM because it works very well with numerical data like ours, handles high-dimensional features, and has been shown in research to give high accuracy for this dataset.

---

4. Teacher: What is a kernel in SVM, and why did you use the RBF kernel?

Student:
A kernel in SVM is a function that helps the model separate data that isn't linearly separable by transforming it into a higher-dimensional space. I used the RBF (Radial Basis Function) kernel because it handles non-linear relationships well, which is common in medical data like ours.

---

5. Teacher: What is label encoding and why is it important?

Student:
Label encoding is the process of converting text labels, like "benign" and "malignant", into numbers, such as 0 and 1. This is important because machine learning models require numerical input, not text, to work properly.

---

6. Teacher: What are hyperparameters? Can you give examples from your project?

Student:
Hyperparameters are settings that I choose before training the model, and they control how the learning process works. They are not learned from the data. For example, in SVM, "C" (cost) controls the trade-off between correct classification and margin width, and "gamma" controls how far the influence of a single training example reaches.

---

7. Teacher: What is a confusion matrix? Can you explain TP, TN, FP, and FN in your context?

Student:
A confusion matrix is a table that shows how many predictions were correct and incorrect.

- True Positive (TP): Model correctly predicts malignant.
- True Negative (TN): Model correctly predicts benign.
- False Positive (FP): Model predicts malignant, but it is actually benign.
- False Negative (FN): Model predicts benign, but it is actually malignant.

---

8. Teacher: What do accuracy, precision, recall, and F1-score mean?

Student:

- **Accuracy** is the proportion of total predictions that are correct.
- **Precision** is the proportion of predicted malignant cases that are actually malignant.
- **Recall** is the proportion of actual malignant cases that the model correctly identified.

- **F1-score** is the harmonic mean of precision and recall, balancing both.

---

9. Teacher: What is an ROC curve and what does AUC represent?

Student:
An ROC curve is a plot that shows how well the model can separate the two classes at different thresholds. The AUC (Area Under the Curve) summarizes this ability—an AUC of 1 means perfect separation, while 0.5 means the model is guessing.

---

10. Teacher: What are SHAP and LIME, and why are they important?

Student:
SHAP and LIME are tools for explaining machine learning model predictions. SHAP shows how much each feature contributed to a prediction, while LIME explains individual predictions using a simple local model. They are important because they help doctors understand why the model made a certain decision, which builds trust and helps in clinical use.

---

11. Teacher: What is supervised learning?

Student:
Supervised learning is a type of machine learning where the model is trained using data that has both inputs and known outputs (labels). The model learns the relationship between the two so it can predict the output for new, unseen data.

---

12. Teacher: Why do we need this model if doctors already do lab tests?

Student:
This model is not meant to replace doctors, but to help them. Even after lab tests, diagnosis can sometimes be unclear or take time. The model can quickly analyze the measurements, provide a second opinion, reduce errors, and help doctors make faster and more confident decisions, especially in difficult or borderline cases.

Certainly! Here are **more viva questions and student-style answers** covering technical, practical, and conceptual aspects of your breast cancer prediction project:

---

13. Teacher: How is the data in the Wisconsin Breast Cancer Dataset collected and calculated?

Student:
The data is collected from patients using a fine needle aspirate (FNA) to take samples from breast lumps. These samples are placed on slides, and high-resolution images are taken under a microscope. Computer software then analyzes the images to calculate measurements like radius, area, texture, and other features for each cell nucleus. These values are then used as features in the dataset.

---

14. Teacher: What is the difference between parameters and hyperparameters?

Student:
Parameters are values that the model learns automatically during training, like the weights in a neural network or the coefficients in regression. Hyperparameters are settings that I choose before training, like the value of "C" and "gamma" in SVM, or the number of trees in a random forest.

---

15. Teacher: Can you explain what overfitting is and how you avoid it in your project?

Student:
Overfitting happens when a model learns the training data too well, including its noise and outliers, and performs poorly on new data. To avoid overfitting, I used only the most important features, split the data into training and test sets, and used cross-validation and hyperparameter tuning.

---

16. Teacher: What is feature scaling and why is it important for SVM?

Student:
Feature scaling means adjusting all features to a similar range, usually by normalization or standardization. It’s important for SVM because it relies on distance calculations—if features are on different scales, the model might give too much importance to features with larger values.

---

17. Teacher: What other algorithms did you compare with SVM, and how did they perform?

Student:
I compared SVM with algorithms like Random Forest, Logistic Regression, Decision Tree, k-Nearest Neighbors, and Artificial Neural Networks. SVM and ANN gave the highest accuracy, usually above 97%, while Logistic Regression and Decision Trees were a bit lower. Random Forest also performed well, but SVM was more consistent.

---

18. Teacher: What is the difference between precision and recall, and why are both important in medical diagnosis?

Student:
Precision tells us, out of all the cases predicted as malignant, how many were actually malignant. Recall tells us, out of all actual malignant cases, how many the model found. Both are important because in medical diagnosis, we want to avoid both false alarms (low precision) and missed cancers (low recall).

---

19. Teacher: What are the limitations of your model and dashboard?

Student:
The main limitation is that the model requires technical measurements from medical tests, so patients can’t use it directly without a doctor’s report. Also, the model’s accuracy depends on the quality and diversity of the dataset. It may not perform as well on data from different populations or imaging equipment.

---

20. Teacher: How does your dashboard help doctors in real life?

Student:
The dashboard lets doctors quickly enter tumor measurements and get an instant prediction, along with explanations and performance metrics. This helps them make faster, more confident decisions, double-check their own diagnosis, and explain results to patients more clearly.

---

21. Teacher: What is AUC, and what does a high AUC mean for your model?

Student:
AUC stands for Area Under the ROC Curve. It shows how well the model can separate benign from

malignant cases at all thresholds. A high AUC (close to 1) means the model is very good at distinguishing between the two classes.

---

22. Teacher: What is the role of the “predict” button in your dashboard?

Student:
The “predict” button takes the values entered by the user for each feature, runs them through the trained SVM model, and displays whether the tumor is likely benign or malignant, along with the explanation and performance metrics.

---

23. Teacher: Can you explain what SHAP and LIME do in simple terms?

Student:
SHAP and LIME are tools that explain why the model made a certain prediction. SHAP shows how much each feature influenced the result, while LIME builds a simple model just for that one prediction to show which features mattered most.

---

24. Teacher: Why is explainability important in medical AI?

Student:
Explainability is important because doctors need to trust and understand the model’s predictions before using them in patient care. If the model can show which features led to its decision, doctors can check if that makes sense and feel more confident using the tool.

---

25. Teacher: What is supervised learning, and how is it different from unsupervised learning?

Student:
Supervised learning uses labeled data, where the correct answers are already known, to train the model. Unsupervised learning uses data without labels and tries to find patterns or groupings on its own. My project uses supervised learning because we know if each tumor is benign or malignant.
