ML Project Midterm Report

1. Introduction/Background

The thyroid is an important part of the human endocrine system, regulating hormones that control processes such as body temperature regulation and blood pressure [2]. Thyroid cancer accounts for around 2.2 percent of new cancer cases and is the most common cancer of the endocrine system [1]. Diagnosis may require blood tests, ultrasound imaging, a biopsy, or in some cases the removal of part of the thyroid [2].

Dataset Details:

The dataset includes 1232 data points and 18 features. These features include blood tests such as FT3, FT4, TSH, TPO, and TGAb [4]. They include data collected from ultrasound and from biopsy. In addition is the feature indicating whether the thyroid nodule was malignant or not, which can be used as a label for supervised learning models.

Dataset: https://zenodo.org/records/6465436

2. Problem Definition

Identifying potential malignant thyroid nodules is a time consuming task for radiologists and is sometimes prone to indeterminate results [1].

We aim to develop and test several Machine Learning algorithms to predict whether thyroid nodules are malignant to reduce invasive procedures.

In an article published in the National Library of Medicine, researchers attempted to utilize several machine learning methods to predict whether a particular thyroid nodule is benign or malignant [1]. Their machine learning methods included:

- Gradient boosting machine
- Logistic regression
- Linear discriminant analysis
- Support vector machine
- Random forest

3. Methods

a. Data Preprocessing Methods

- **Model 1: KMeans**
  - **Data Transformation:** We will normalize and standardize the dataset to ensure uniform feature scales. Normalization adjusts values to a common range (e.g. 0 to 1), critical for models like neural networks, while standardization ensures a mean of zero and a standard deviation of one, improving model performance.
  - **Feature Selection:** The features we selected—shape, calcification, echo pattern, size, composition, and margin—are widely recognized as significant indicators for predicting thyroid cancer. Each feature has been chosen for its established clinical relevance in assessing thyroid nodules and its ability to contribute to meaningful clustering in our analysis. Below is an

overview of each feature and its role in predicting malignancy, supported by research findings.

Shape is a categorical feature with values of 0 for regular shapes and 1 for irregular shapes. Nodules with an irregular or "taller-than-wide" shape are often associated with a higher malignancy risk, as this orientation suggests vertical growth, which is more typical of malignant nodules. While binary and limited in variance, this feature has clinical value due to its direct correlation with malignancy [5].

Calcification indicates whether calcium deposits are present within the nodule (0 = absent, 1 = present). Microcalcifications are a strong predictor of thyroid cancer, frequently observed in malignant nodules. Despite being binary, calcification's high correlation with malignancy underscores its importance as a predictive feature [6].

Echo pattern is another binary feature (0 = even, 1 = uneven) that describes the thyroid's echogenicity or texture uniformity. Hypoechoic (darker) nodules with uneven echo patterns are more likely to be malignant, as this reduced echogenicity is characteristic of cancerous nodules. While binary, this feature holds significant predictive power due to its association with malignancy [7].

Unlike previous features, size is a continuous variable, measuring the physical dimensions of the nodule in centimeters. Larger nodules generally carry a higher malignancy risk, with nodules over 2 cm warranting closer evaluation. Size's continuous nature provides valuable variance that enhances cluster separation, making it a valuable feature for clustering. However, if the range of sizes is narrow, the model's predictions may be limited to nodules within this range [6].

Composition indicates whether the nodule is cystic (0), mixed (1), or solid (2). Solid nodules have a higher risk of malignancy compared to cystic or mixed nodules. Although categorical with limited variance, composition helps in distinguishing malignancy types, as solid nodules are generally more concerning [8].

The margin feature describes the boundary clarity of the nodule (0 = clear, 1 = unclear). Nodules with irregular or unclear margins are more likely to be malignant, as these margins suggest invasive growth patterns. Despite being binary, margin clarity is a critical predictor of malignancy and a valuable feature in clustering [7].

- **Model 2: Neural Network**
    - **Data Transformation:** Neural Networks require input features to be on a similar scale for the best training, making normalization and standardization really important. Continuous variables like size are normalized between 0 and 1. This ensures all features contribute equally during gradient-based optimization. Standardization is also applied to ensure the mean of all features is zero and the standard deviation is one. This prevents biases arising from feature magnitude differences and accelerates convergence during backpropagation [9].
    - **Dimensionality Reduction:** To reduce the computational complexity and focus on the most significant patterns, we apply Principal Component Analysis (PCA) to the dataset. PCA extracts 11 components, representing the majority of the variance in the data. By transforming the input features into this lower-dimensional space, we reduce redundancy and enhance the neural network's ability to learn critical relationships [10].
    - **Feature Selection and Extraction:** While the selected clinical features (shape, calcification, echo pattern, size, composition, and margin) guide initial preprocessing, PCA ensures that only the most informative aspects of these features are retained for model training.
- **Model 3: Random Forest**

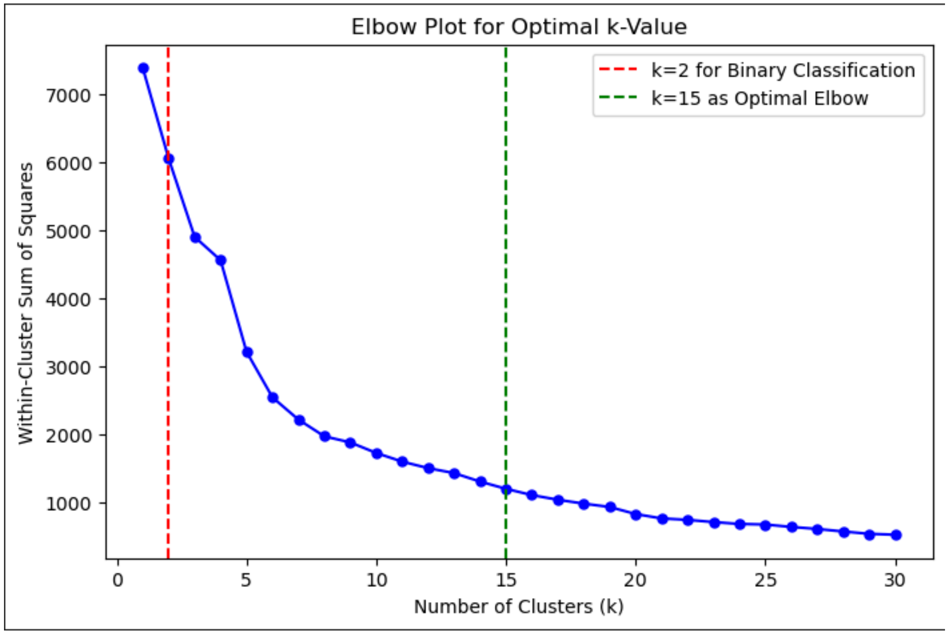
  - **Data Transformation:** The dataset undergoes standardization, ensuring a mean of zero and a standard deviation of one for all features. This preprocessing step supports interpretability in feature importance rankings and smoothes compatibility with metrics like SHAP values [11]. Additionally, normalization is skipped, as the Random Forest algorithm does not rely on distance metrics sensitive to feature magnitude.
  - **Feature Selection:** The same features selected for K-Means—shape, calcification, echo pattern, size, composition, and margin—are employed for Random Forest, as these features hold strong clinical significance. Random Forest naturally incorporates feature selection during training by measuring the importance of features based on splits at each decision node. This process ensures that irrelevant or redundant features are down-weighted automatically, reinforcing the clinical relevance of these selected features [12].
  - **Feature Importance and Extraction:** To further analyze the feature contributions, we visualize the top 10 most important features as determined by the Random Forest's intrinsic feature importance metric. This step provides actionable information into how strongly each feature contributes to the prediction task. By specifying critical predictors like calcification or shape, we ensure our model prioritizes features with the highest clinical relevance.
  - **Data Splitting:** To evaluate the generalization of the Random Forest model, the dataset is divided into training (80%) and test (20%) sets. A cross-validation procedure is also applied during hyperparameter tuning to prevent overfitting and assess model stability across different splits.
- **Feature Selection and Extraction:** We will remove irrelevant or redundant features, reducing model complexity, preventing overfitting, and speeding up computation. Techniques such as correlation analysis or tree-based selection will help us identify key predictors, like critical blood test results.
- **Data Splitting:** For our neural network, the dataset will be divided into training and test sets to evaluate model generalization. A validation set may also be used to fine-tune hyperparameters and prevent overfitting.

b. ML Algorithms/Models

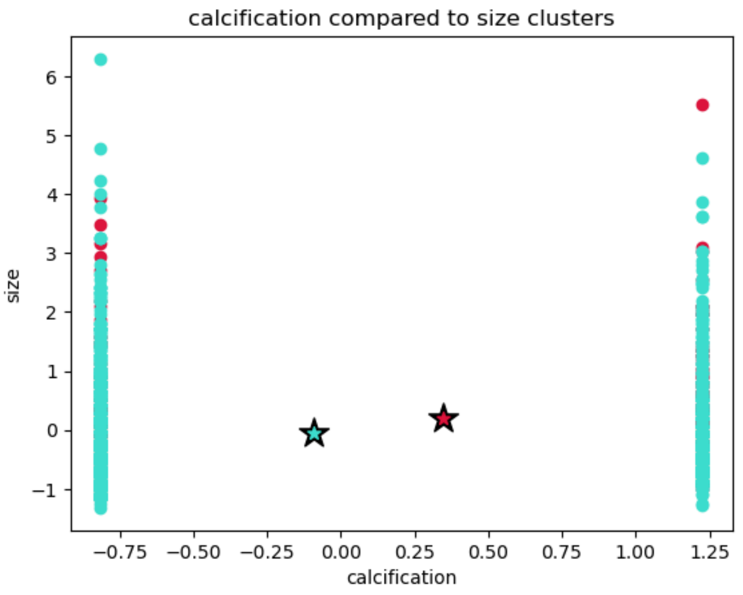
- **Model 1: K-Means:** This unsupervised algorithm clusters data points based on feature similarity, uncovering patterns or subtypes in thyroid nodule data that could inform later predictions.
  - **K-Value Selection:** For our analysis, we chose a k-value of 2 in the K-Means clustering algorithm to align with our goal of binary classification to distinguish between benign and malignant thyroid nodules. By setting k=2, we can create two distinct clusters that intuitively represent these two categories, providing a straightforward way to assess whether nodules are likely benign or malignant based on their cluster assignments. This choice simplifies the interpretation of clustering results by directly supporting our objective of evaluating malignancy risk.

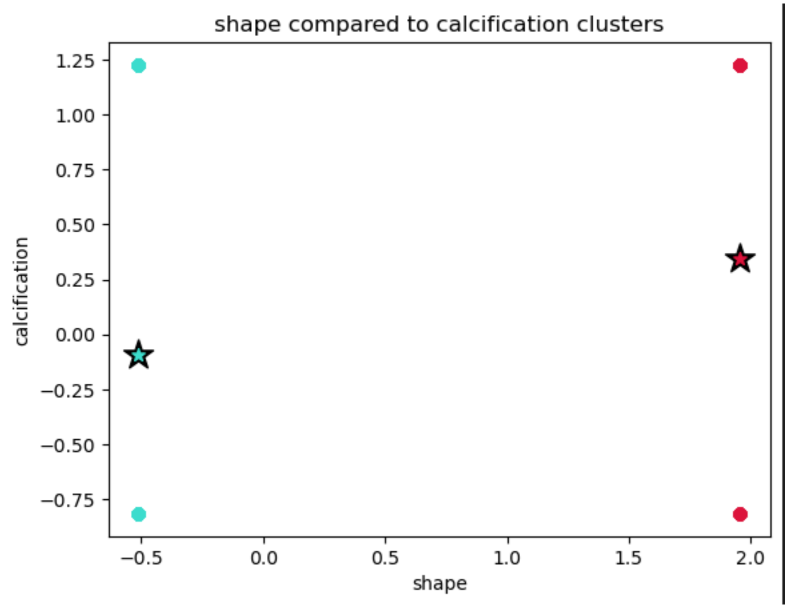
    While our elbow graph suggests that k=15 may be optimal from a clustering perspective (as it minimizes the within-cluster sum of squares effectively), selecting k=2 enhances interpretability within the binary classification framework. Using k=15 could reveal further granular patterns, but it would fragment the data into smaller groups that don’t directly support the primary focus of this project.

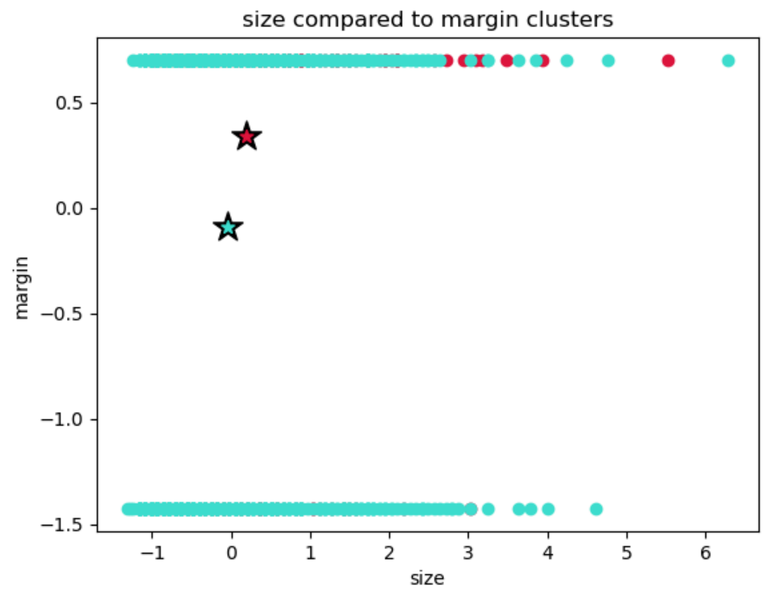
    Thus, choosing k=2 enables a more clinically relevant interpretation, grouping nodules into two main clusters that can be examined for malignancy tendencies. This approach allows us to leverage clustering for binary classification, directly addressing the question of whether a nodule is more likely to be benign or malignant.

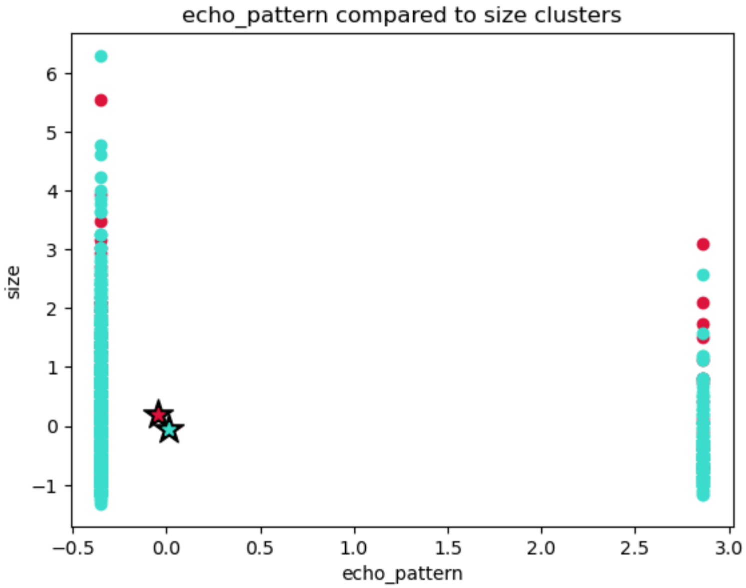


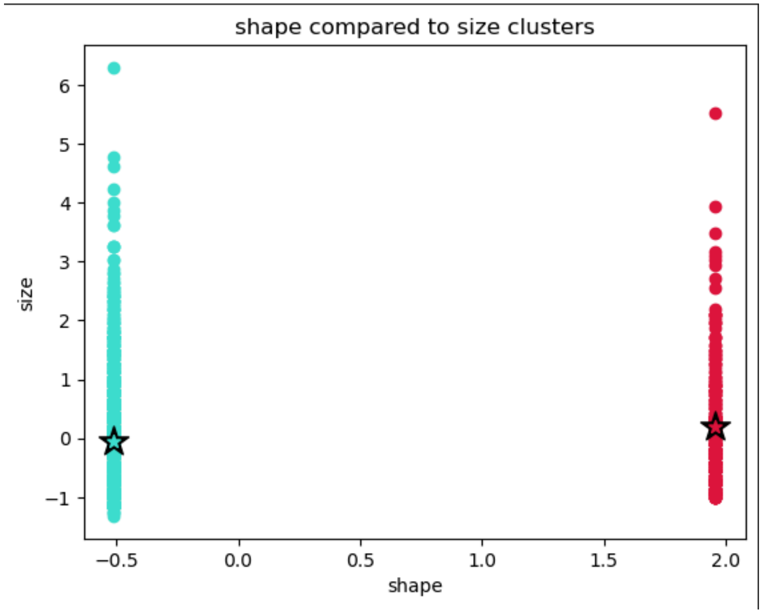
- **2D and 3D Clustering Plots:**

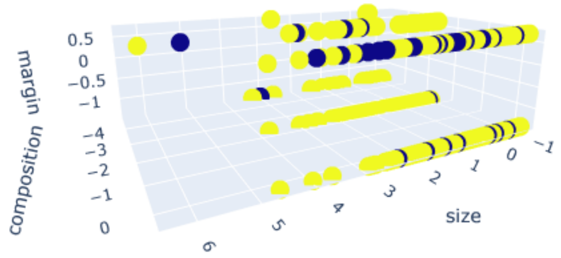


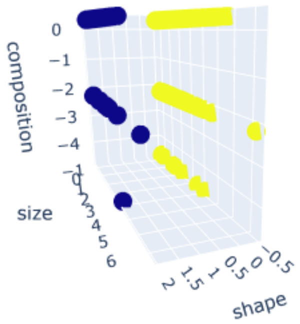

shape compared to calcification clusters
calcification
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−0.5
0.0
0.5
1.0
1.5
2.0
shape


size compared to margin clusters
margin
0.5
0.0
−0.5
−1.0
−1.5
−1
0
1
2
3
4
5
6
size


echo_pattern compared to size clusters
size
6
5
4
3
2
1
0
−1
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
echo_pattern


shape compared to size clusters
size
6
5
4
3
2
1
0
−1
−0.5
0.0
0.5
1.0
1.5
2.0
shape





- **Discrete Features:** Our dataset includes some discrete features, such as “calcification” and “composition,” which are binary or categorical rather than continuous. While discrete features are generally less ideal for K-Means clustering, as K-Means relies on calculating distances that work better with continuous data, we retained certain discrete features that are clinically relevant for thyroid cancer prediction. For instance, features like “calcification” and “echo pattern” provide important diagnostic information, so removing them could reduce the model’s predictive accuracy. Although these features may be harder to visualize and analyze with K-Means, their inclusion helps ensure the model captures critical clinical information relevant to malignancy prediction.

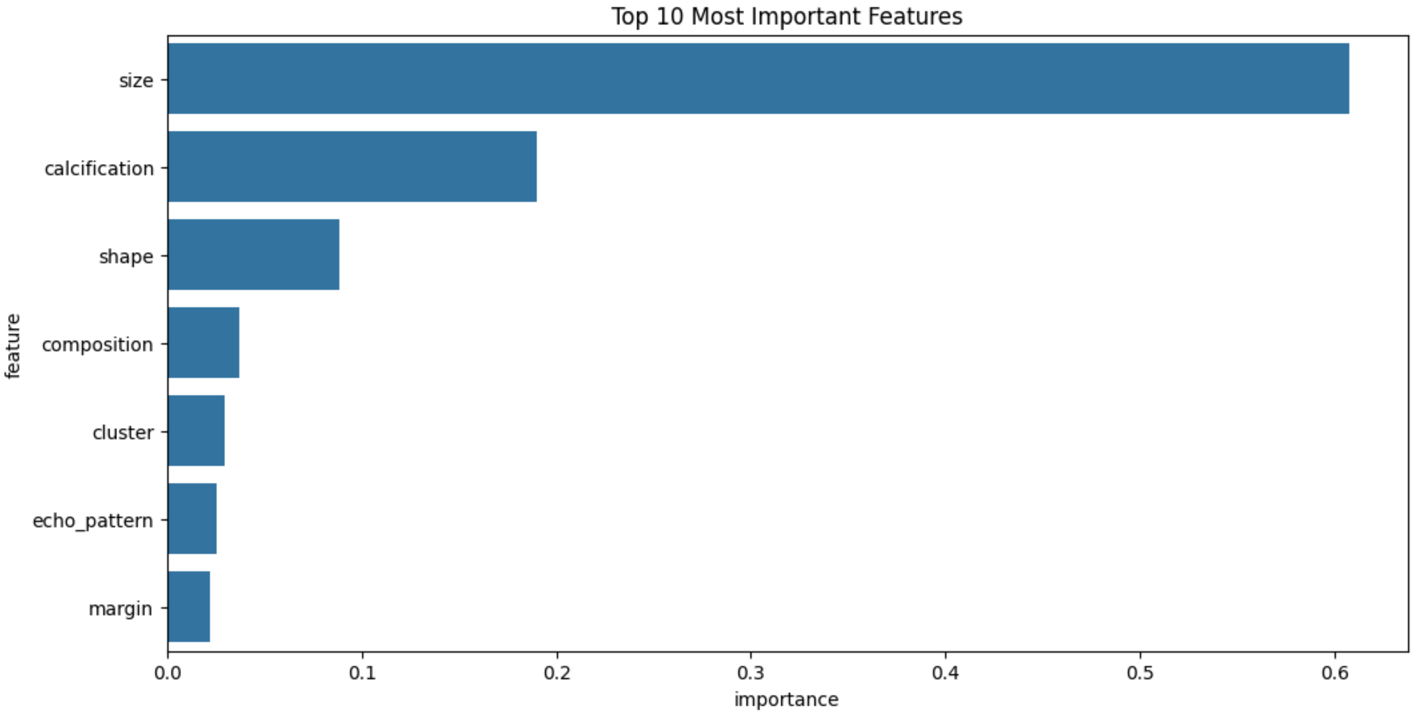
- **Model 2: Neural Network:** The Neural Network is a supervised learning model that identifies complex patterns in the dataset to classify thyroid nodules as benign or malignant. A neural network will capture complex patterns for malignancy prediction, prioritizing recall to minimize false negatives. Techniques like regularization will help prevent overfitting.
  - **Architecture:** Input layer = 11 features (after PCA). Hidden layers = Two layers with 11 neurons each. Output layer = Logistic activation for binary output.

  - Its flexibility in modeling nonlinear relationships makes it well-suited for this task. The neural network architecture consists of an input layer with 11 features obtained from Principal Component Analysis (PCA). These features represent the most important patterns in the data while reducing redundancy. The model also includes two hidden layers, each with 11 neurons, using logistic activation functions to handle nonlinear transformations. The output layer consists of a single neuron with a logistic activation function to generate binary predictions.

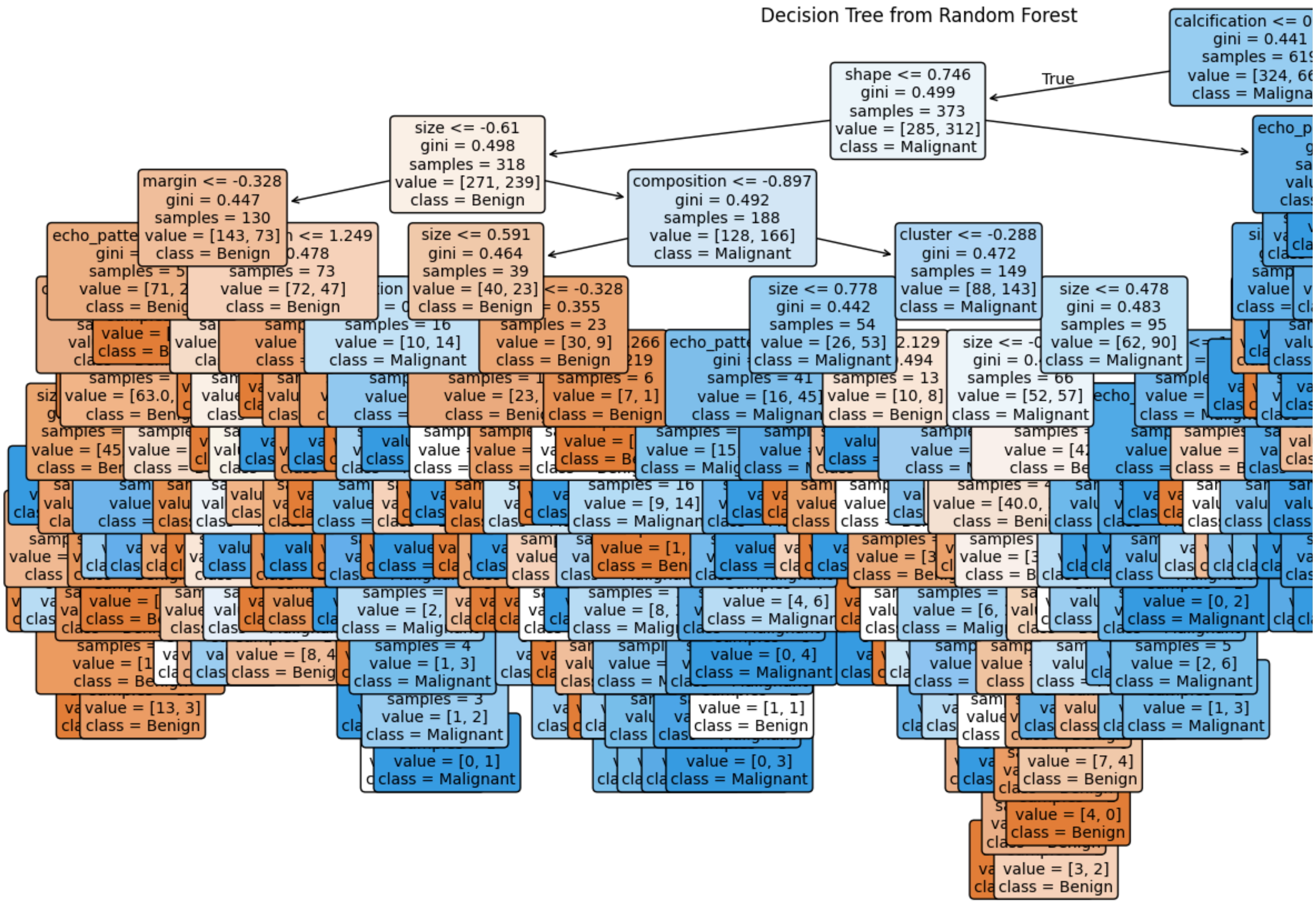
  - To optimize performance, the neural network employs regularization techniques like L2 regularization (α=0.05\alpha = 0.05α=0.05) to prevent overfitting by discouraging large weights. The Adam optimizer is chosen because of its adaptive learning rates, which help the model converge faster and more reliably.
  - The model's performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Among these, recall is prioritized since minimizing false negatives is critical in cancer prediction to ensure malignant cases are not missed. Visualizations like loss curves can be used to show how the training loss decreases over iterations, validating the model’s convergence. Additional graphs such as the ROC curve and precision-recall curve help illustrate the trade-offs between sensitivity, specificity, and positive prediction power. A confusion matrix heatmap provides a clear view of true positives, true negatives, false positives, and false negatives to highlight areas for improvement.
  - While PCA simplifies the input features, some clinical interpretability may be lost due to dimensionality reduction. To address this, it is important to explain how the original features contribute to the principal components. Furthermore, handling any class imbalance in the dataset could improve recall performance, and techniques like oversampling or weighting the classes could be explored.

- **Model 3: Random Forest:** This supervised model uses multiple decision trees to classify data, with feature importance ranking and majority voting making it well-suited for predicting thyroid nodule malignancy.
  - The Random Forest is another supervised model used for this project. It combines multiple decision trees to classify thyroid nodules as benign or malignant. By using an ensemble of trees, Random Forest achieves robust predictions that are less prone to overfitting compared to single decision trees. Each tree is trained on a random subset of the data, and predictions are made based on majority voting across all the trees. This ensemble method makes Random Forest particularly effective for complex datasets like this one.
  - One of the key strengths of Random Forest is its ability to rank features based on their importance. Features like calcification, shape, and size are expected to rank highly because they are clinically significant for thyroid cancer prediction. Random Forest calculates feature importance by measuring how much each feature reduces impurity during splits in the decision trees. A bar chart showing the top-ranked features can provide valuable insights into which features contribute most to classification.
  - The model is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation ensures the model generalizes well to new data. To further enhance interpretability, SHAP (SHapley Additive exPlanations) values are used to explain how individual features influence predictions. Visualizations such as the confusion matrix help identify misclassifications, while an ROC curve illustrates the trade-offs between sensitivity and specificity. A visual representation of one of the decision trees from the ensemble offers additional interpretability by showing the decision-making process in the model.
  - While Random Forest is robust to multicollinearity, examining correlated features could still improve interpretability and efficiency. Additionally, hyperparameter tuning, such as optimizing the number of trees or tree depth, can further balance performance and computational efficiency. Visualizing SHAP values with a summary plot or using a feature importance bar chart can help better understand the model’s decision-making process.

- **Feature Importance Classification:**



- **Decision Tree:**



4. (Potential) Results and Discussion

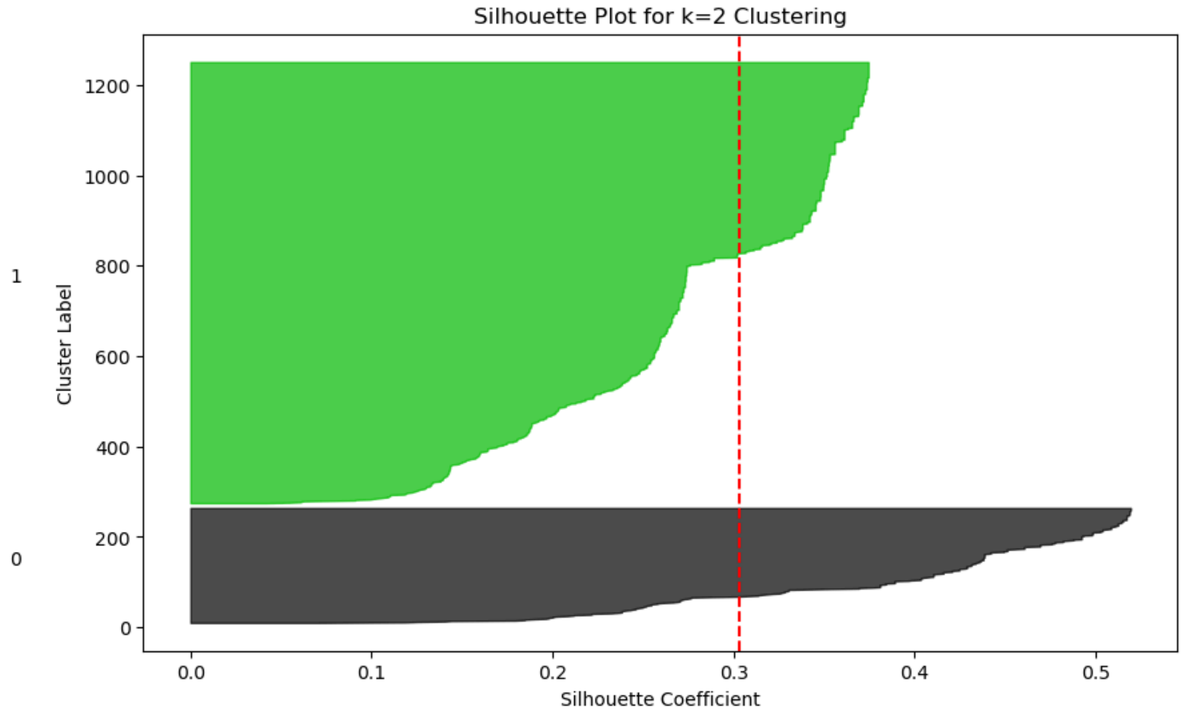
Model 1: K-Means:

- **Accuracy and Confusion Matrix** To evaluate the effectiveness of our K-Means clustering in predicting thyroid malignancy, we used three quantitative metrics: Adjusted Rand Score (ARS), Silhouette Score, and clustering accuracy (derived from the confusion matrix). Each of these metrics offers a unique perspective on clustering performance, highlighting areas of both strength and limitation.

- 1. Adjusted Rand Score (ARI)

  The Adjusted Rand Score (ARI) for our clustering results was calculated as **-0.0392**. ARI measures the similarity between the clusters formed by K-Means and the true labels, adjusted for chance. An ARI score close to 1 would indicate strong alignment, while an ARI near 0 suggests minimal alignment. With a low ARI score of -0.039212, our clustering model shows limited alignment with the true labels, implying that the separation into two clusters does not fully capture the underlying structure of benign and malignant nodules. This is expected, as K-Means is an unsupervised algorithm and does not leverage label information when creating clusters.
- 2. Silhouette Score

  The Silhouette Score for our model was 0.303. This score, which ranges from -1 to 1, measures the cohesion and separation of clusters. A score closer to 1 indicates well-defined, cohesive clusters, while a score near 0 suggests overlapping clusters. Our Silhouette Score of 0.303 suggests moderate clustering quality, indicating that the data supports a two-cluster division but with some overlap between clusters. This moderate score aligns with the clinical complexity of thyroid cancer prediction, as benign and malignant cases can share similar characteristics.

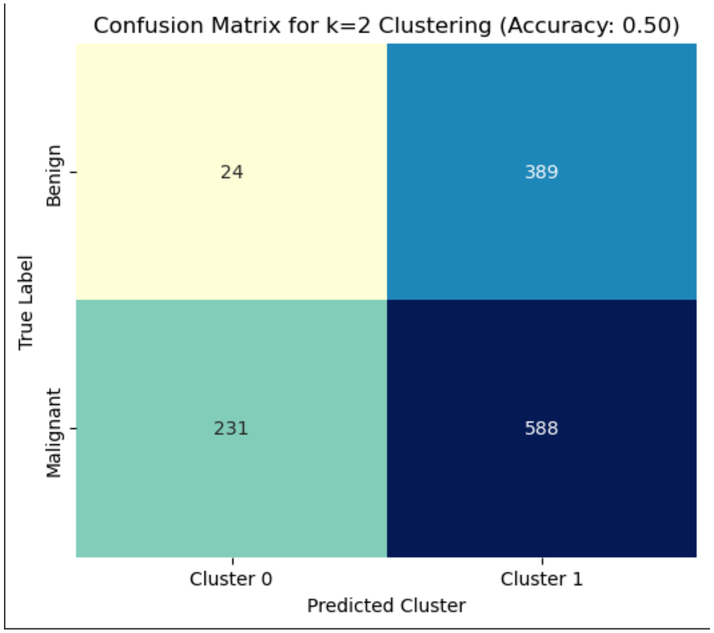


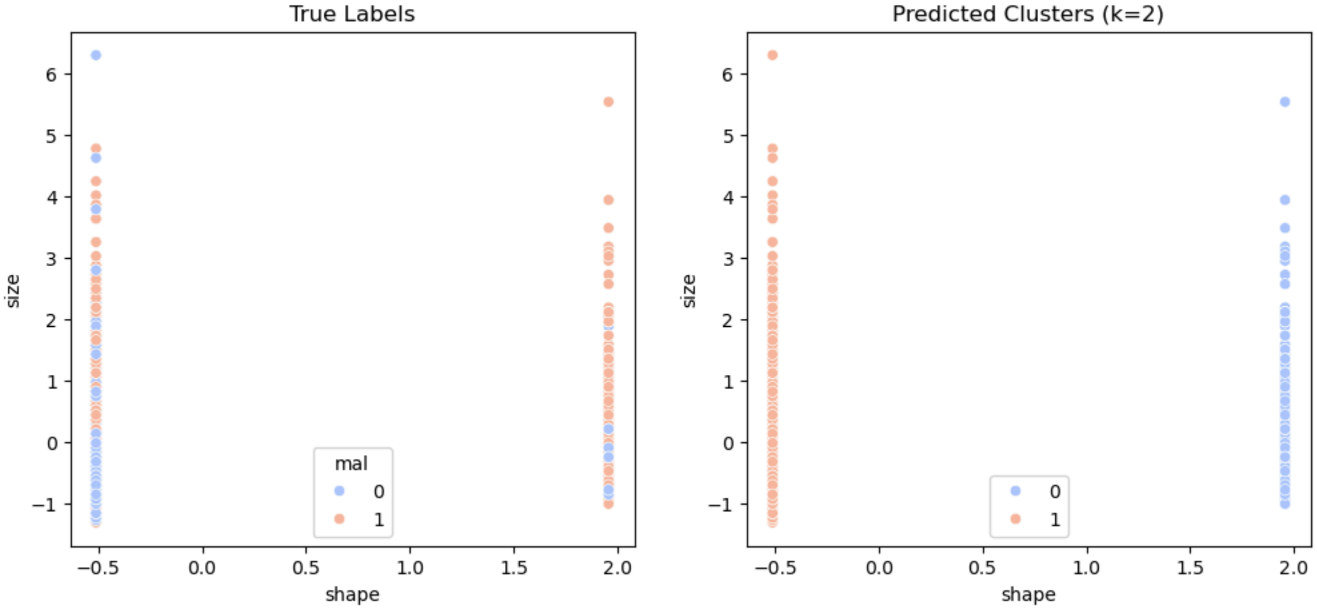
- 3. Accuracy and Confusion Matrix

  The confusion matrix provides insights into the clustering accuracy by comparing the clusters with true labels:

  Correctly Clustered Instances: 612 instances were correctly clustered (24 benign and 588 malignant).

  Misclustered Instances: 620 instances were incorrectly assigned (389 benign cases misclustered as malignant and 231 malignant cases misclustered as benign).





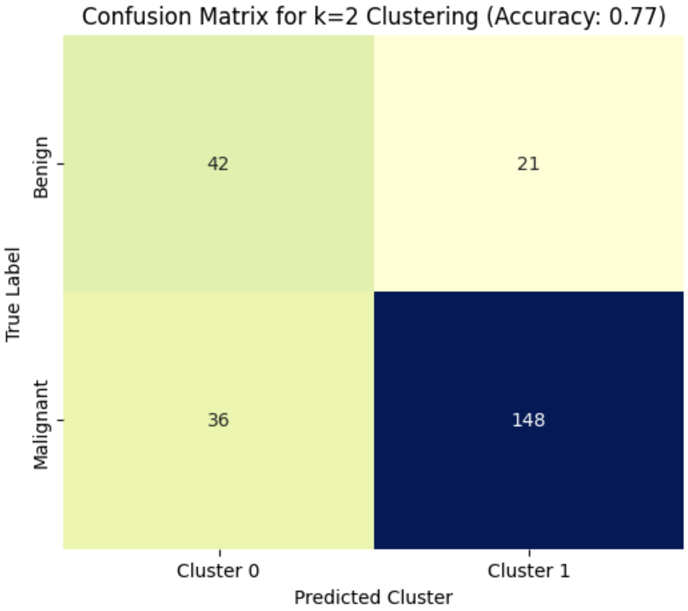
Overall, this clustering yielded an approximate accuracy of: Accuracy = (24 + 588)/(24 + 389 + 231 + 588) = 0.4868. This result suggests that the model correctly classified only about 49% of the cases. The confusion matrix further reveals that our model struggled particularly with benign nodules, as 389 benign cases were misclassified as malignant, while the model performed relatively better with malignant cases. This could indicate that features associated with benign and malignant cases may overlap, affecting the cluster purity for benign nodules.

- **Summary:** Our K-Means clustering model demonstrates some success in distinguishing benign and malignant thyroid nodules, achieving an approximate accuracy of 49%. It performs well in identifying malignant cases, correctly clustering 588 instances, suggesting that our selected features—shape, calcification, echo pattern, size, composition, and margin—capture meaningful distinctions. The Silhouette Score of 0.303 further indicates that these clinically relevant features help form reasonably defined clusters. Setting k=2 aligns with our binary classification goal, clearly grouping nodules into high-risk or low-risk categories, supporting effective initial screening. This analysis provides a strong foundation for refining accuracy through supervised learning methods in the future.

We believe that our models may perform similarly to the ones described in section 2 if not better in some cases [1]. We plan to use our unsupervised machine learning method to classify and understand our data, while our supervised methods predict whether a thyroid nodule is malignant with high accuracy.

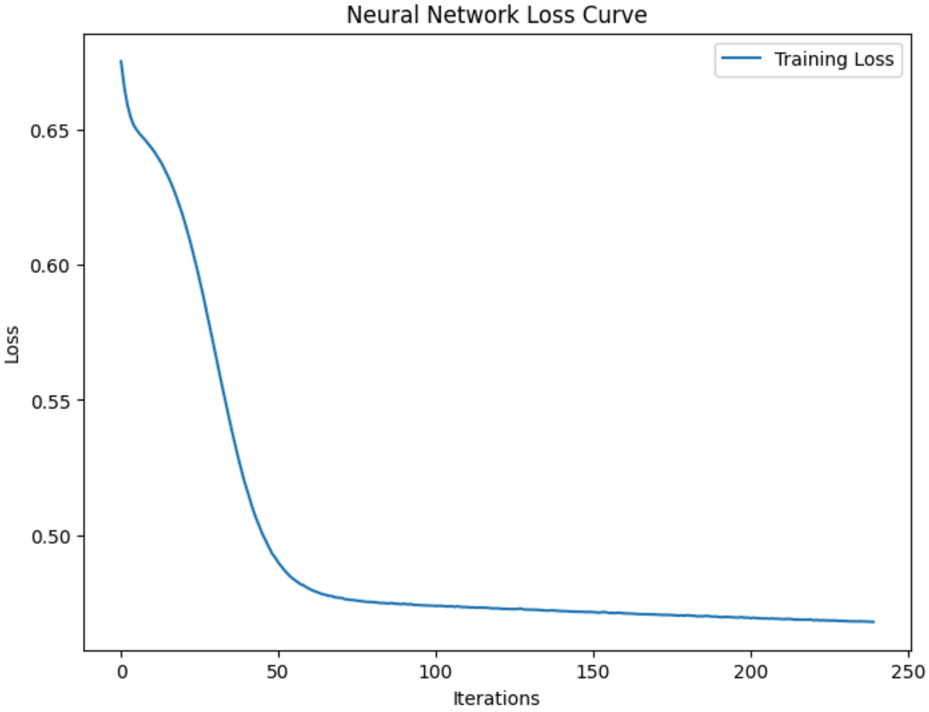
Model 2: Neural Network:

- **Accuracy and Confusion Matrix:**
    - To evaluate the performance of our neural network in predicting thyroid malignancy, we analyzed its accuracy using the confusion matrix, along with other metrics like the ROC-AUC score. The confusion matrix revealed that the model achieved an accuracy of 0.77. This indicates that the model correctly classified 77% of thyroid nodules as either benign or malignant. While this is a promising result, it also suggests room for improvement, as 23% of the nodules were misclassified. The confusion matrix further shows that the model performed better in identifying malignant cases compared to benign ones, reflecting the clinical challenge of distinguishing between these categories due to overlapping features.

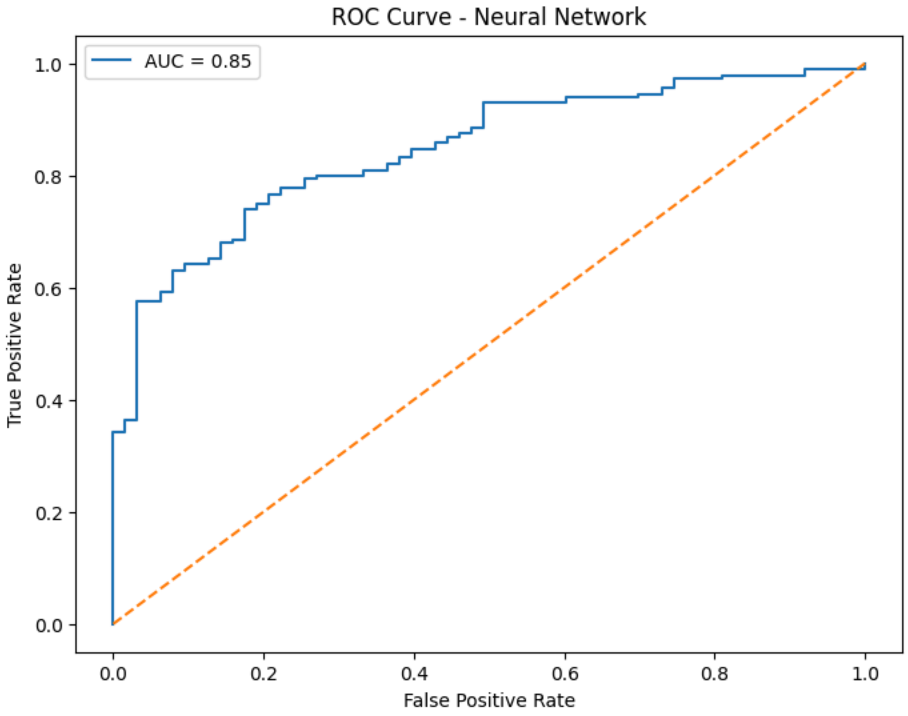


- **Loss Curve:**
    - The loss curve demonstrates how the model's training loss decreased over iterations. Initially, the loss was 0.65, but it dropped rapidly to 0.5 within the first 50 iterations, after which it plateaued for the remaining 200 iterations. This indicates that the model quickly learned significant patterns in the data but struggled to refine its predictions further. The plateau

suggests that the model may have reached a local minimum or encountered limitations in learning due to the complexity of the data or model architecture.



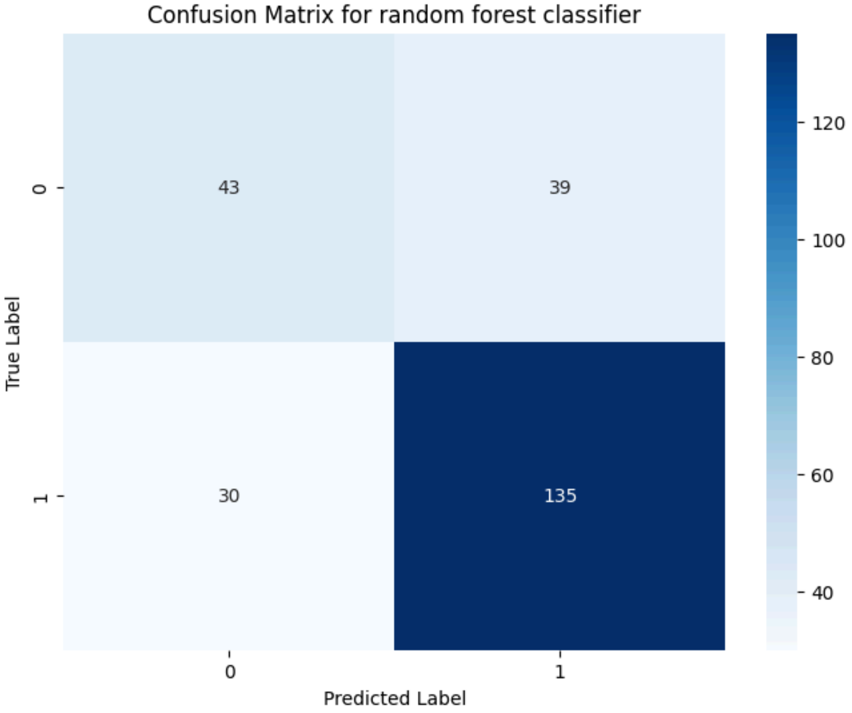
- **ROC Curve:**
  - The ROC curve provides insights into the model's ability to differentiate between benign and malignant nodules. The AUC score of 0.85 indicates strong discriminatory power, suggesting that the model performs well overall in distinguishing between the two classes. The curve shows a favorable balance between sensitivity and specificity, reinforcing the model’s capability in identifying malignant nodules with high confidence while minimizing false negatives, which is crucial in thyroid cancer diagnosis.



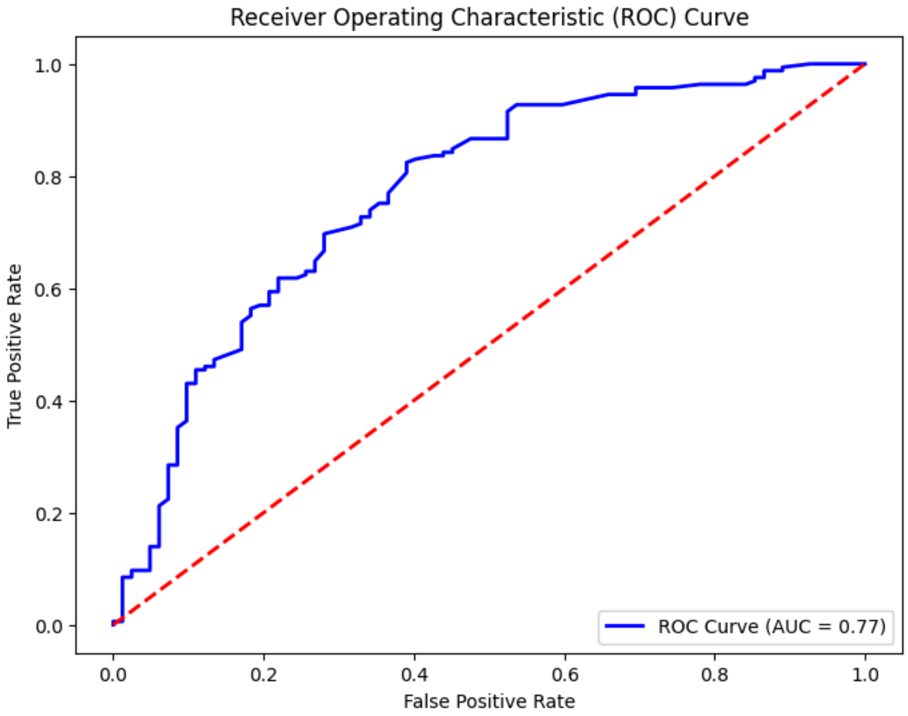
- **Summary:** The neural network demonstrates strong performance in predicting thyroid malignancy, with an accuracy of 77% and an AUC of 0.85. The rapid drop in the loss curve highlights the model's ability to learn from the data efficiently, though its plateau suggests the need for potential refinements, such as hyperparameter tuning or architectural adjustments. These results indicate that the neural network is effective at capturing complex patterns in the dataset, making it a reliable tool for binary classification. However, there is potential to improve performance by addressing class imbalances, incorporating additional features, or employing advanced techniques like ensemble learning.

Model 3: Random Forest:

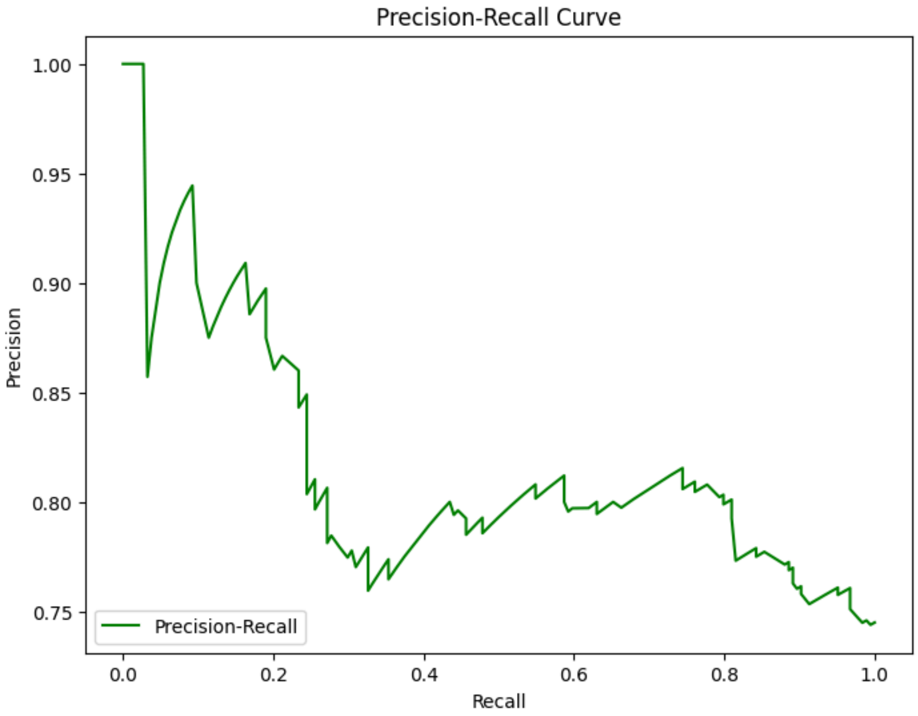
- **Accuracy and Confusion Matrix:**
    - The Random Forest model achieved an accuracy of 0.72, as shown by the confusion matrix. This indicates that 72% of thyroid nodules were correctly classified as benign or malignant. While this result suggests reasonable predictive performance, the model's ability to distinguish between classes could still be improved. A closer examination of the confusion matrix reveals that the model performs better in identifying malignant cases compared to benign ones. This imbalance could reflect the clinical challenge of differentiating between these categories due to overlapping feature distributions.



- **ROC Curve:**
    - The ROC curve further evaluates the model's performance by plotting the trade-off between the true positive rate (sensitivity) and the false positive rate across various thresholds. The model achieved an AUC of 0.77, indicating moderate discriminatory power. While not as high as desired, this score suggests that the model is effective at distinguishing between benign and malignant nodules in most cases, though there is room for optimization to enhance sensitivity and reduce false negatives.



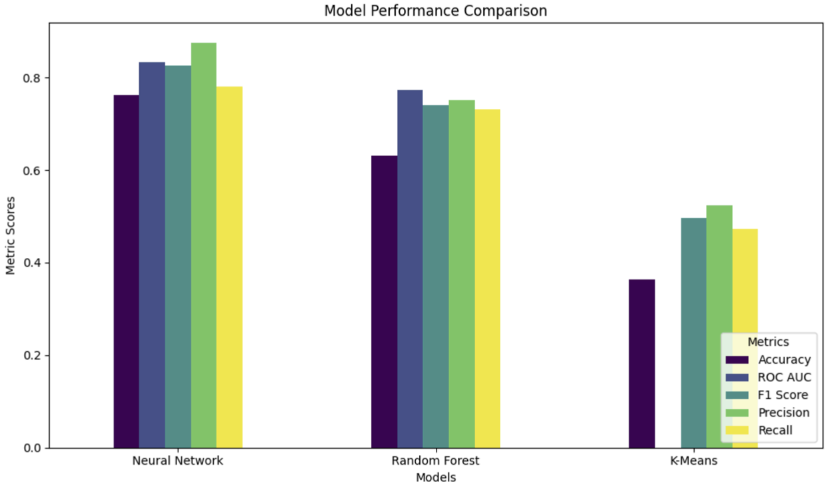
- **Precision-Recall Curve:**
    - The precision-recall curve provides additional insights into the model’s performance in identifying true positives, particularly for imbalanced datasets. As seen in the curve, precision starts at 1.0 when recall is low, indicating perfect predictions for a small subset of cases. However, precision steadily declines as recall increases, eventually dropping below 0.75 at high recall levels. This trend highlights a trade-off: while the model achieves high precision for specific thresholds, maintaining that precision while improving recall (identifying more malignant cases) is a challenge. Despite this, the model demonstrates reasonably consistent precision above 0.75 for much of the curve, reflecting reliable performance for many thresholds.



- **Feature Importance and Decision Tree Analysis:**
    - The feature importance graph revealed that calcification, shape, and margin were the most influential predictors in the model. These features are consistent with known clinical indicators of thyroid malignancy, suggesting that the model is utilizing meaningful data to make predictions. Additionally, the decision tree visualization offered interpretability into the model's logic, showing how splits on key features contribute to the overall prediction process.

- **Summary:** The Random Forest model displayed moderate overall performance, with an accuracy of 0.72, an AUC of 0.77, and a balanced precision-recall trade-off. While the model effectively utilizes clinically relevant features like calcification and shape, its performance could be improved by reducing false negatives, especially in identifying malignant cases. Strategies such as hyperparameter tuning, incorporating additional features, or combining Random Forest with other ensemble methods could help enhance its sensitivity and overall accuracy. These results suggest that the Random Forest model is a reliable tool for thyroid malignancy prediction, though additional refinements could make it more effective in clinical applications.

5. Comparison



- **Performance:**
    - RF and NN are closely matched in terms of accuracy, ROC AUC, F1 Score, precision, and recall.
    - NN slightly edges out RF in recall, making it more sensitive in identifying true positive cases. However, RF’s precision makes it less prone to false alarms.
- **Interpretability:**
    - RF stands out for its ability to provide feature importance, making it more actionable for real-world applications.
    - NN, despite its strong performance, lacks interpretability, making it less suitable for decision-critical applications without additional explainability techniques.
- **Scalability and Efficiency:**
    - K-Means is computationally efficient and useful for exploratory data analysis but performs poorly on supervised classification metrics.
    - RF and NN require more computational resources but offer much better predictive performance.
- **K-Means**
    - **Strengths:**

        - While K-Means operates as an unsupervised algorithm, its clusters can provide valuable insights into the underlying structure of the data, such as separating benign and malignant cases.
        - K-Means is computationally inexpensive and simple to implement.
    - **Limitations:**
        - The unsupervised nature of K-Means makes it less effective at aligning clusters with true labels. This explains its poor performance across supervised metrics.
    - **Use Case:**
        - Best used as a preliminary tool to explore patterns in data. Its utility is limited in classification tasks compared to supervised models like NN and RF.
    - **Recommentations:**
        - Utilize K-Means for exploratory purposes to uncover data structure.
        - Do not rely on K-Means for tasks requiring high classification accuracy.
- **Neural Network:**
    - **Strengths:**
        - The NN model shows a high ROC AUC score, indicating strong ability to distinguish between classes. This suggests the model is effective in ranking positive cases higher than negatives.
        - Balanced and consistently high scores across these metrics demonstrate that the NN model is good at handling the trade-off between precision (minimizing false positives) and recall (minimizing false negatives).
    - **Limitations:**
        - While NN excels in predictive accuracy, its performance may vary significantly with hyperparameter tuning and feature scaling. This could be computationally expensive and require fine-tuning.
        - Its interpretability is limited compared to RF.
    - **Use Case:**
        - Suitable for scenarios where model interpretability is not critical, and the focus is on maximizing predictive performance.
    - **Recommentations:**
        - Leverage NN when the priority is maximizing predictive performance and computational resources are available.
        - Combine NN with explainability tools like SHAP or LIME for better interpretation.
- **Random Forest:**
    - **Strengths:**
        - RF performs slightly better than NN in terms of accuracy and F1 score, indicating it is highly effective at identifying both benign and malignant cases.
        - RF has the highest precision among the models, making it the most reliable in minimizing false positives.
        - RF provides insights into feature relevance, making it more interpretable and actionable.
    - **Limitations:**
        - Slightly lower recall compared to NN indicates a slight trade-off in identifying all true positives.
        - May struggle with overfitting if not properly regularized, but this can be mitigated with hyperparameter tuning.
    - **Use Case:**
        - Ideal for clinical settings where interpretability and reliability are essential. The ability to identify important features can help clinicians understand the key drivers of predictions.
    - **Recommentations:**
        - Use RF in clinical applications where interpretability is key.
        - Focus on tuning hyperparameters (e.g., number of estimators, max depth) to further enhance performance.

6. References

[1] D. L. Cooper et al., "COVID-19 Vaccines and Autoimmune Diseases," National Center for Biotechnology Information (NCBI), 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9249901/#CR1

[2] S. K. Mohanty et al., "Thyroid Dysfunction and Its Relation to Metabolic Syndrome," National Center for Biotechnology Information (NCBI), 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415174/

[3] A. P. Weetman, "The Immunopathogenesis of Chronic Autoimmune Thyroiditis," Thyroid, vol. 20, no. 8, pp. 823–826, 2010. [Online]. Available: https://www.liebertpub.com/doi/full/10.1089/thy.2009.0455

[4] "Thyroid Function Tests," American Thyroid Association (ATA), 2023. [Online]. Available: https://www.thyroid.org/thyroid-function-tests/

[5] Oxford Academic, "Thyroid nodule malignancy risks based on shape orientation," [Online]. Available: https://academic.oup.com/jcem/article/107/7/1865/6570820?login=false

[6] Thyroid Foundation, "Patient thyroid information: Indicators of thyroid cancer," [Online]. Available: https://www.thyroid.org/patient-thyroid-information/ct-for-patients/may-2018/vol-11-issue-5-p-8-9/

[7] Springer Link, "Thyroid imaging and echogenicity in malignancy prediction," [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-44100-9_10

[8] Radiologic Clinics, "Thyroid nodule composition and malignancy correlations," [Online]. Available: https://www.radiologic.theclinics.com/article/S0033-8389(19)30001-6/abstract

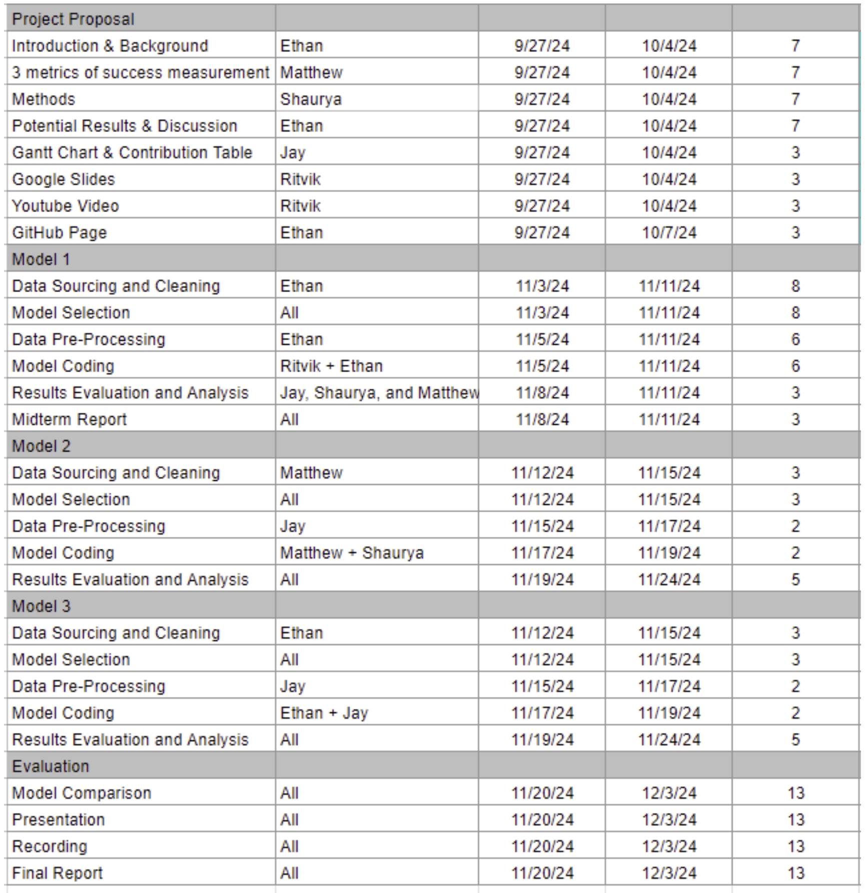
[9] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning and feature scaling," [Online]. Available: https://www.deeplearningbook.org/

[10] H. Abdi and L. J. Williams, "Principal component analysis: Concepts, methodology, and applications," WIREs Comput Stat, [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/wics.101

[11] L. Breiman, "Random forests: A classification and regression technique," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[12] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://www.jmlr.org/papers/v12/pedregosa11a.html

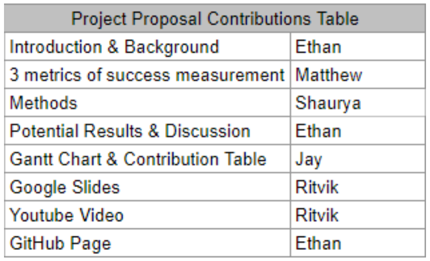
7. Gantt Chart



Project Proposal				
Introduction & Background	Ethan	9/27/24	10/4/24	7
3 metrics of success measurement	Matthew	9/27/24	10/4/24	7
Methods	Shaurya	9/27/24	10/4/24	7
Potential Results & Discussion	Ethan	9/27/24	10/4/24	7
Gantt Chart & Contribution Table	Jay	9/27/24	10/4/24	3
Google Slides	Ritvik	9/27/24	10/4/24	3
Youtube Video	Ritvik	9/27/24	10/4/24	3
GitHub Page	Ethan	9/27/24	10/7/24	3
Model 1				
Data Sourcing and Cleaning	Ethan	11/3/24	11/11/24	8
Model Selection	All	11/3/24	11/11/24	8
Data Pre-Processing	Ethan	11/5/24	11/11/24	6
Model Coding	Ritvik + Ethan	11/5/24	11/11/24	6
Results Evaluation and Analysis	Jay, Shaurya, and Matthew	11/8/24	11/11/24	3
Midterm Report	All	11/8/24	11/11/24	3
Model 2				
Data Sourcing and Cleaning	Matthew	11/12/24	11/15/24	3
Model Selection	All	11/12/24	11/15/24	3
Data Pre-Processing	Jay	11/15/24	11/17/24	2
Model Coding	Matthew + Shaurya	11/17/24	11/19/24	2
Results Evaluation and Analysis	All	11/19/24	11/24/24	5
Model 3				
Data Sourcing and Cleaning	Ethan	11/12/24	11/15/24	3
Model Selection	All	11/12/24	11/15/24	3
Data Pre-Processing	Jay	11/15/24	11/17/24	2
Model Coding	Ethan + Jay	11/17/24	11/19/24	2
Results Evaluation and Analysis	All	11/19/24	11/24/24	5
Evaluation				
Model Comparison	All	11/20/24	12/3/24	13
Presentation	All	11/20/24	12/3/24	13
Recording	All	11/20/24	12/3/24	13
Final Report	All	11/20/24	12/3/24	13

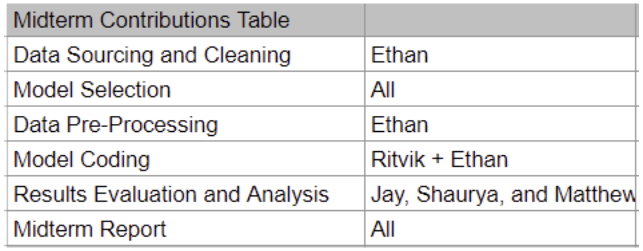
8. Contributions Tables

1. Project Proposal:



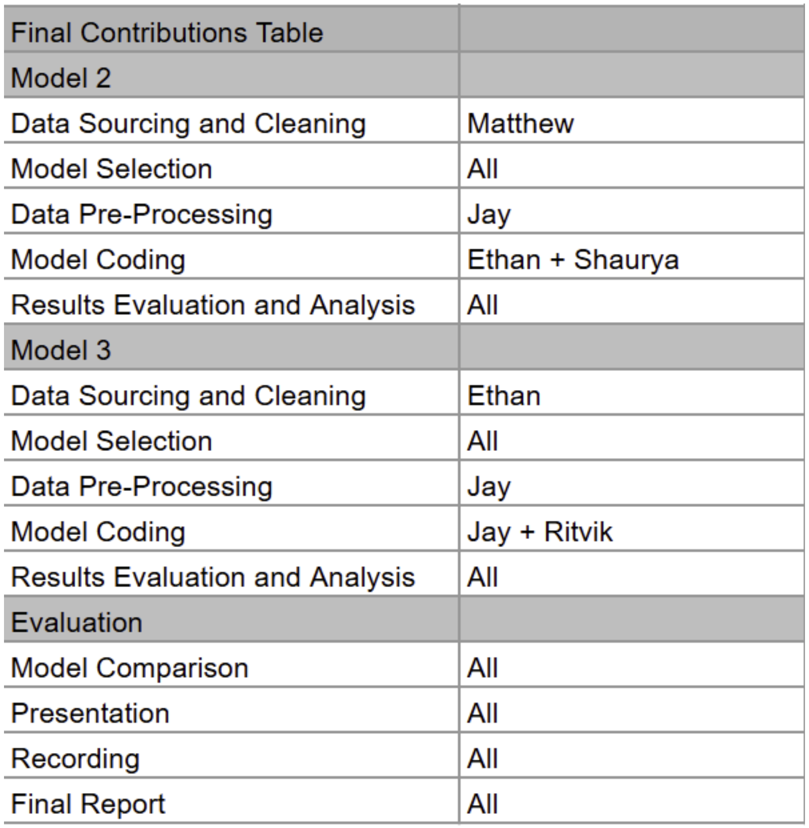
Project Proposal Contributions Table	
Introduction & Background	Ethan
3 metrics of success measurement	Matthew
Methods	Shaurya
Potential Results & Discussion	Ethan
Gantt Chart & Contribution Table	Jay
Google Slides	Ritvik
Youtube Video	Ritvik
GitHub Page	Ethan

2. Project Midterm:



Midterm Contributions Table	
Data Sourcing and Cleaning	Ethan
Model Selection	All
Data Pre-Processing	Ethan
Model Coding	Ritvik + Ethan
Results Evaluation and Analysis	Jay, Shaurya, and Matthew
Midterm Report	All

3. Project Final:



Final Contributions Table	
Model 2	
Data Sourcing and Cleaning	Matthew
Model Selection	All
Data Pre-Processing	Jay
Model Coding	Ethan + Shaurya
Results Evaluation and Analysis	All
Model 3	
Data Sourcing and Cleaning	Ethan
Model Selection	All
Data Pre-Processing	Jay
Model Coding	Jay + Ritvik
Results Evaluation and Analysis	All
Evaluation	
Model Comparison	All
Presentation	All
Recording	All
Final Report	All