TECH TITANS

Exploring innovations that will shape our world tomorrow.

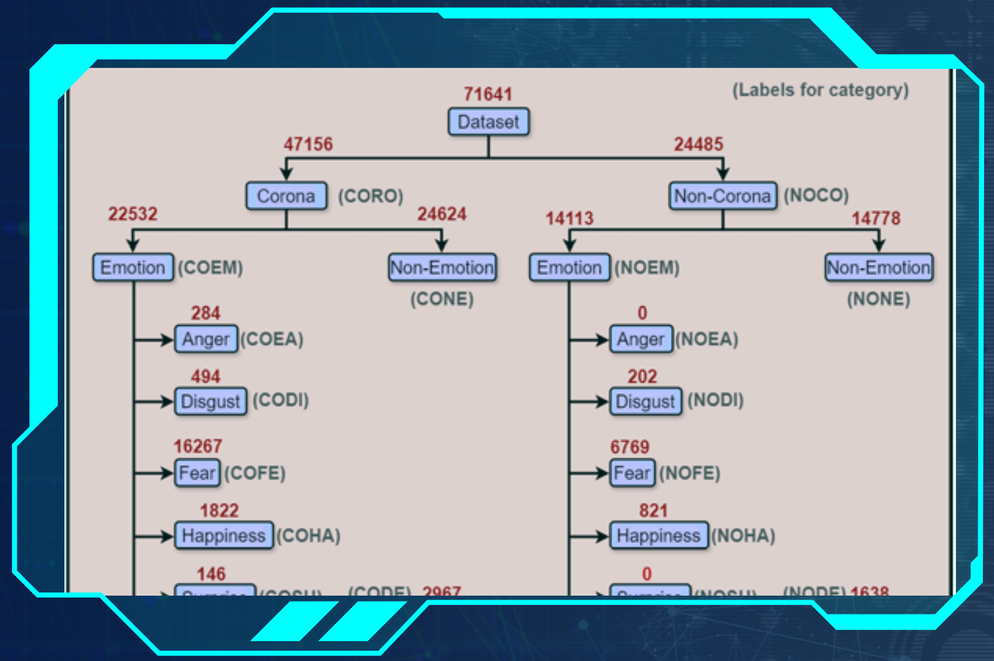
Start

TEAM MEMBERS

1. N CHARAN SAI - RA2311003011343
2. AKASH SINGH - RA2311026010638

ROUND 2

MISSION



Social media platforms like Twitter played a crucial role in public discourse during the COVID-19 pandemic, capturing a wide spectrum of emotions and concerns. However, analyzing these tweets effectively requires a multi-level classification framework to distinguish between COVID-19 and non-COVID-19 discussions, identify emotional content, and further categorize sentiments.

STEP-BY-STEP PROCESS FOR MULTI-LEVEL CLASSIFICATION

1. Define the Problem
2. Data Collection
3. Data Preprocessing
4. Split the Data
5. Model Selection
6. Training the Model
7. Evaluation
8. Fine-Tuning
9. Testing
10. Deployment
11. Monitoring and Maintenance

BERT (BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS)

Developed by: Google

Purpose: Designed to understand word context in sentences by considering both preceding and following words, improving language understanding.

Key Features:

- Contextual Understanding: Reads text bidirectionally (both left-to-right and right-to-left) for better context comprehension.
- Pre-training & Fine-tuning: Pre-trained on large text corpora, fine-tuned for tasks like sentiment analysis and question answering.
- Transfer Learning: Efficiently adapts to various tasks with minimal labeled data.

USE OF TRANSFORMER

We are leveraging pretrained transformer models (like BERT) for hierarchical multi-level classification of tweets.

1. Text Tokenization → Convert tweets into tokenized inputs using BERT tokenizer.
2. Pretrained Model → Use a pretrained transformer (BERT, RoBERTa, DistilBERT) instead of traditional machine learning classifiers.
3. Fine-tuning on Labeled Data → Train the model for multi-level classification.
4. Multi-Level Predictions → Hierarchical classification with different levels.

CHALLENGE WHILE DOING EMBEDDING

1. **Long computation time**

Processing long tweets (max token length = 512 in BERT) consumes more memory.

2. **Handling Informal Language, Sarcasm, and Emojis in Tweets**

Tweets often contain slang, misspellings, emojis, and informal abbreviations

CHALLENGE WHILE DOING EMBEDDING

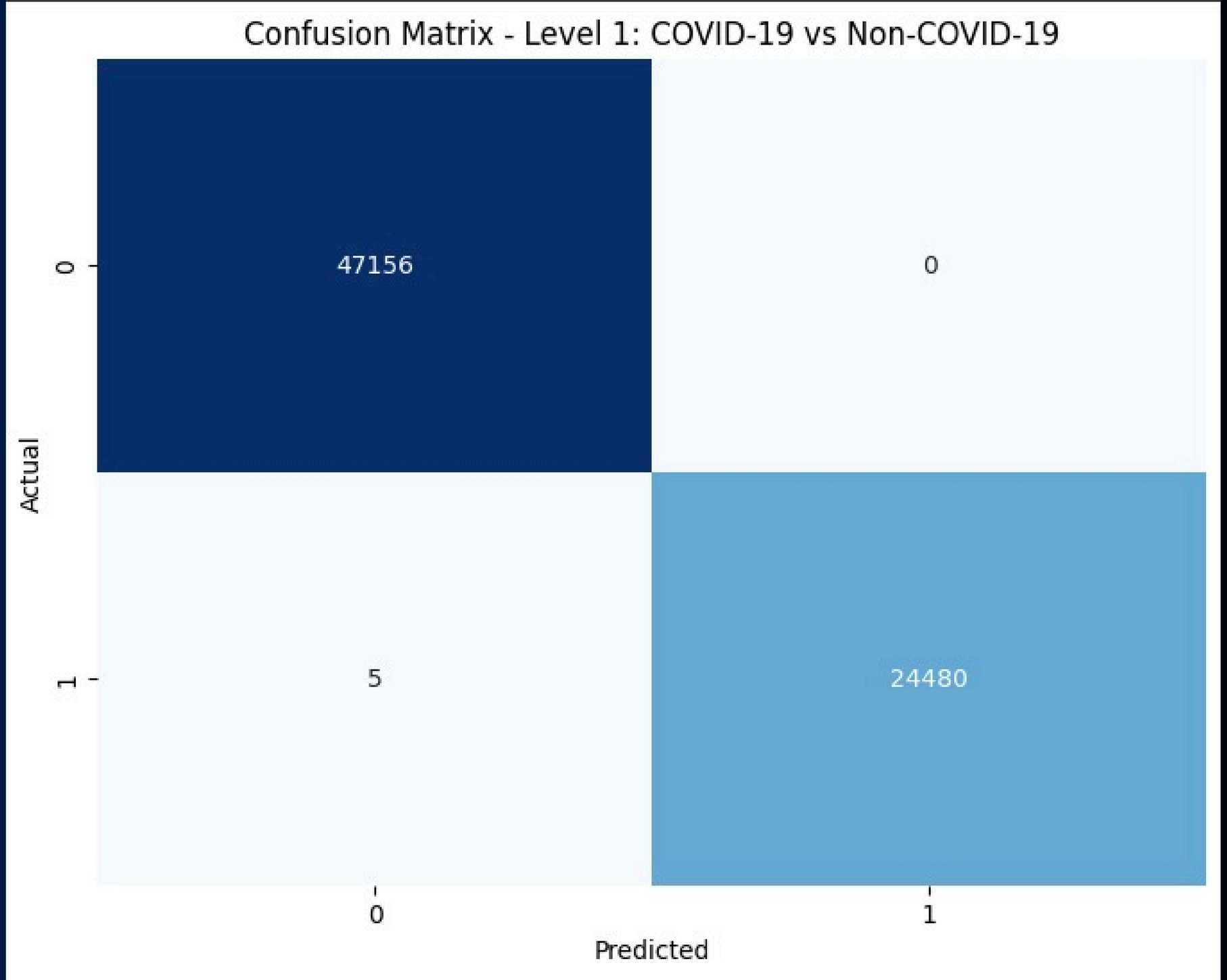
3. Handling Imbalanced Data in Multi-Level Classification

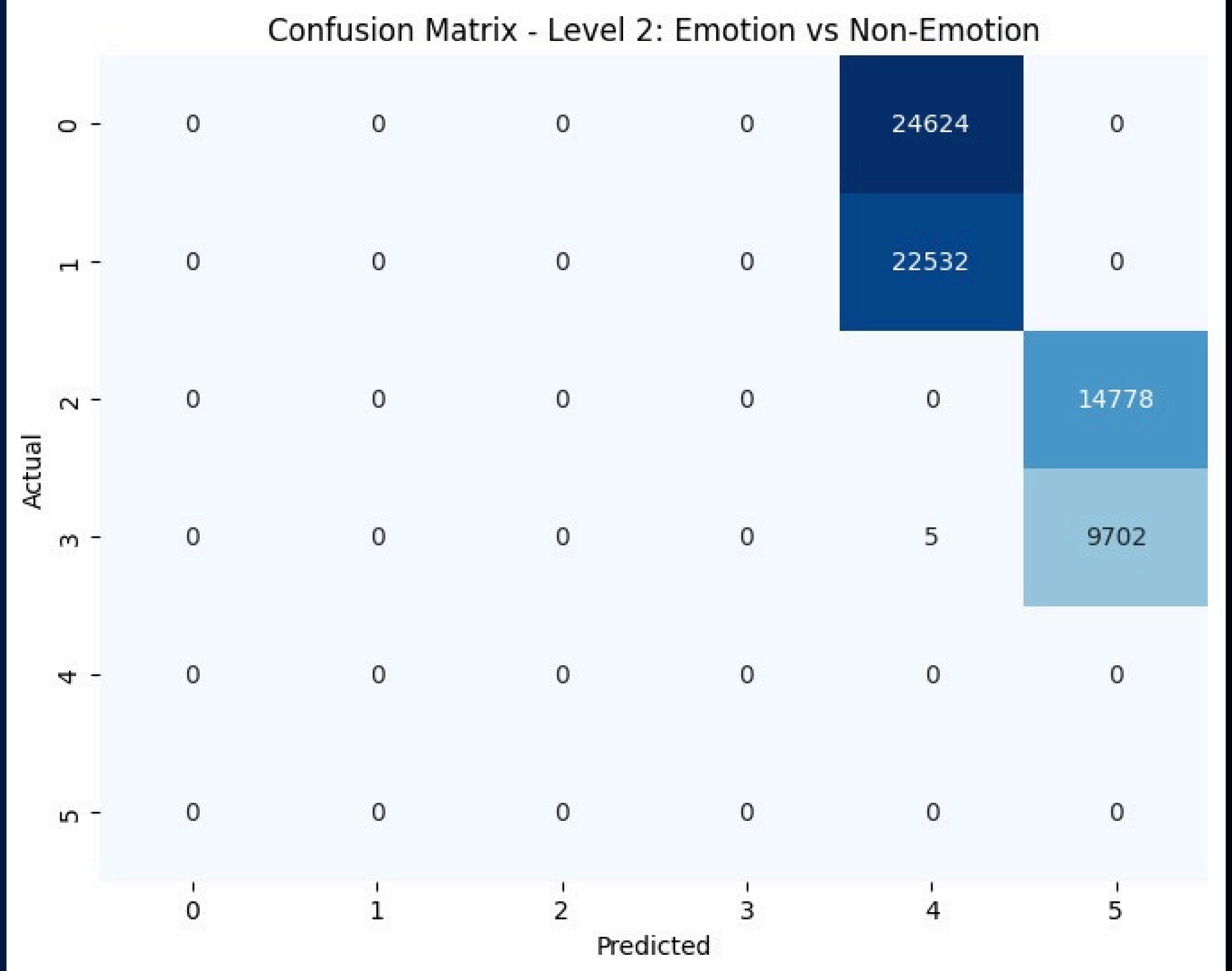
```
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=5e-5),
              loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
              metrics=['accuracy'])

model.fit(train_dataset, epochs=1)  # Adjust epochs as needed
```

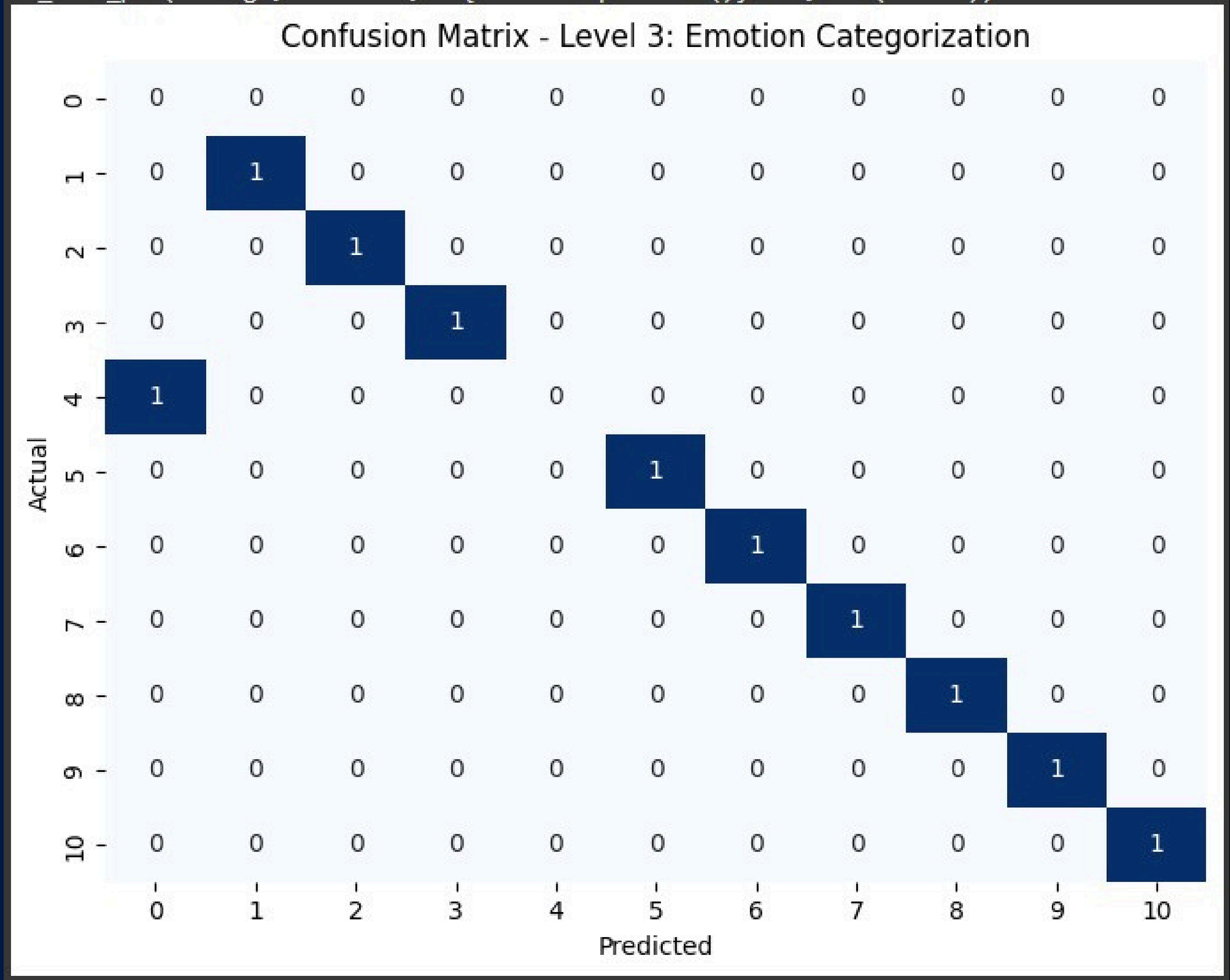
```
2954/2954 [==============================] - 504s 159ms/step - loss: 0.0017 - accuracy: 0.9999
<tf_keras.src.callbacks.History at 0x7f6911eb63d0>
```

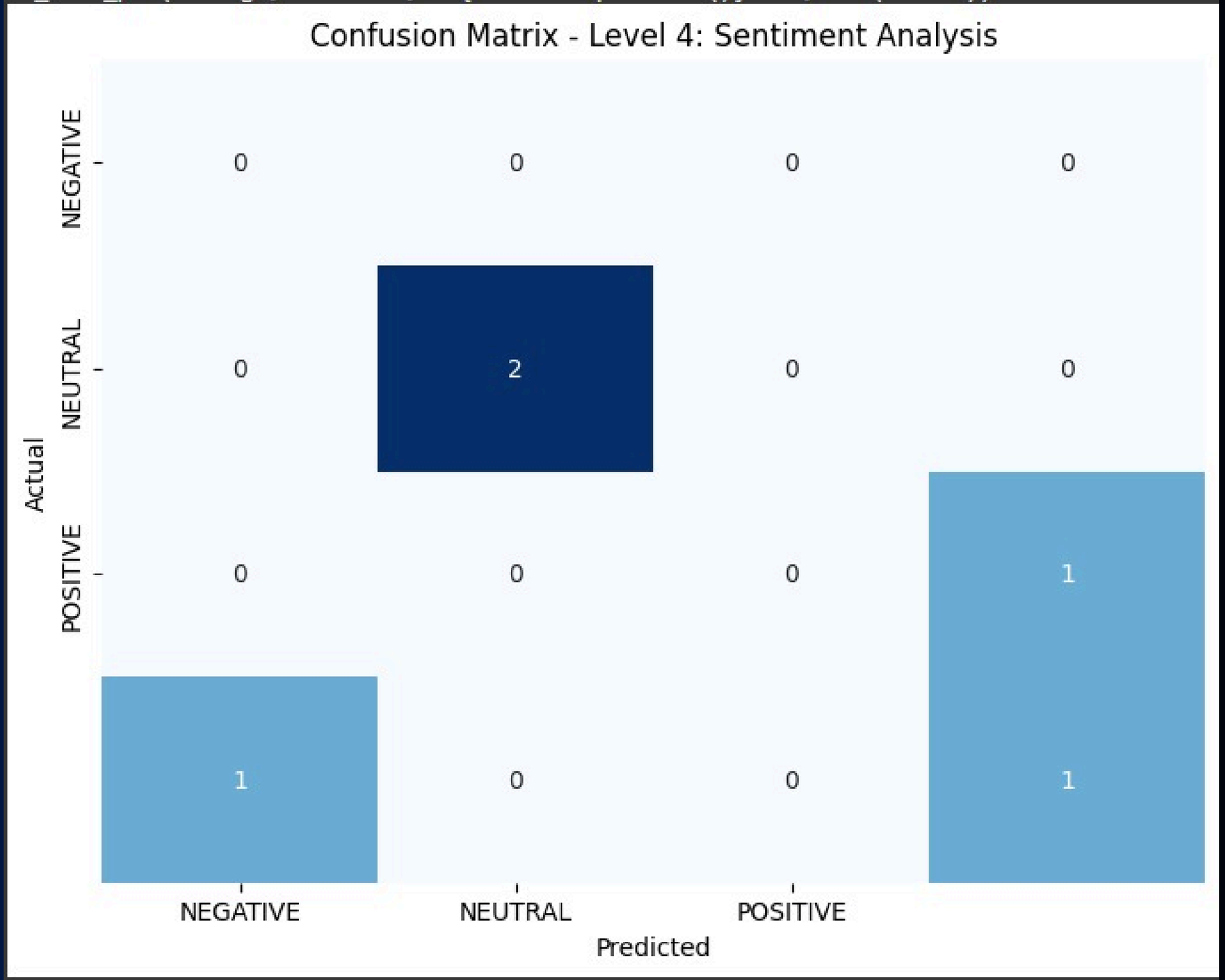
```
predictions = model.predict(test_dataset)
predicted_labels = tf.argmax(predictions.logits, axis=1).numpy()
```

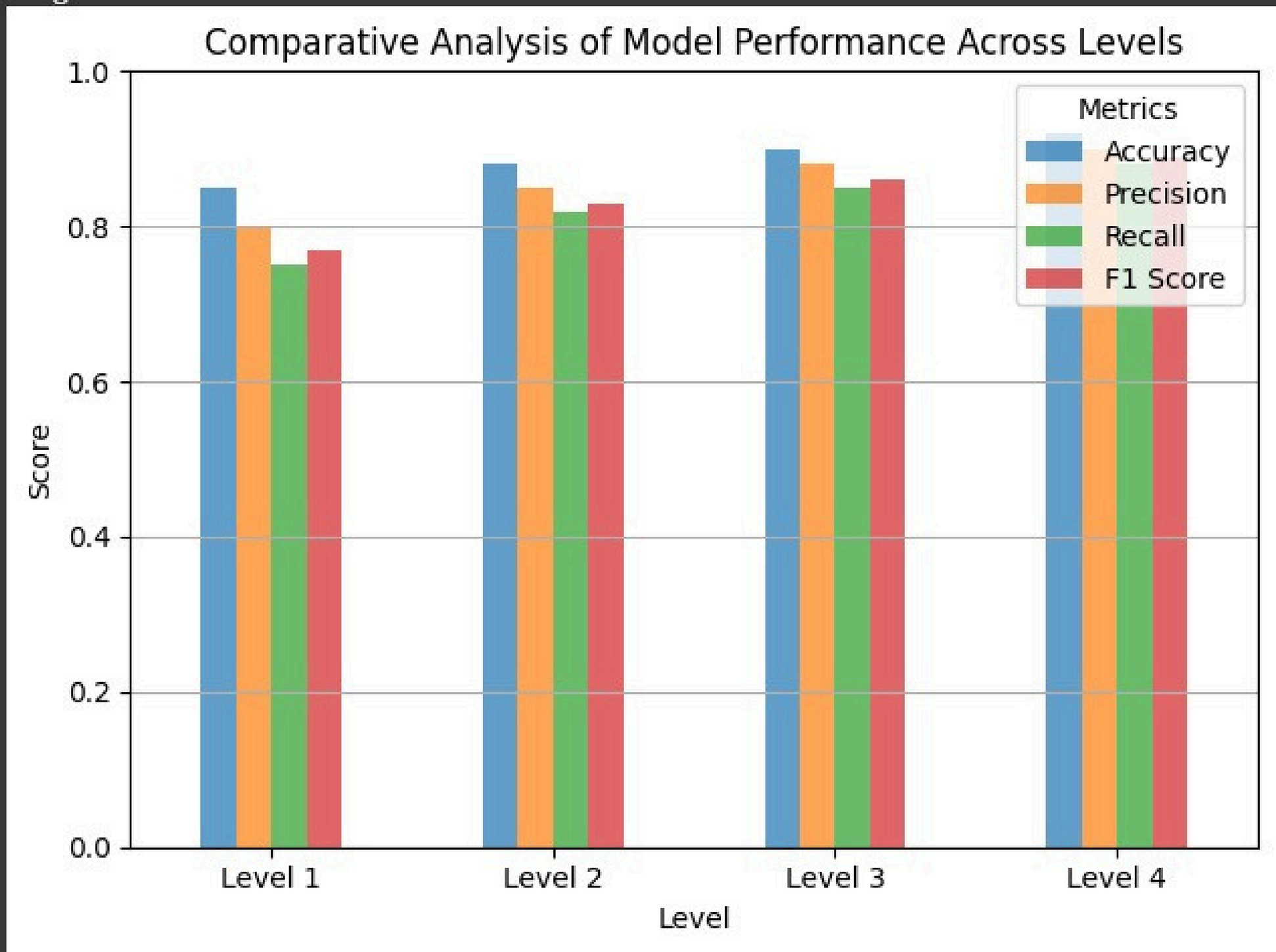
```
739/739 [==============================] - 36s 47ms/step
```


Confusion Matrix - Level 1: COVID-19 vs Non-COVID-19
47156
0
5
24480
Actual
0
1
0
1
Predicted


Confusion Matrix - Level 2: Emotion vs Non-Emotion
Actual
Predicted
0
1
2
3
4
5
0 0 0 0 24624 0
1 0 0 0 22532 0
2 0 0 0 0 14778
3 0 0 0 5 9702
4 0 0 0 0 0 0
5 0 0 0 0 0 0


Confusion Matrix - Level 3: Emotion Categorization
Actual
Predicted
0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 1
0 1 2 3 4 5 6 7 8 9 10


Confusion Matrix - Level 4: Sentiment Analysis
NEGATIVE
NEUTRAL
POSITIVE
Actual
0
0
0
0
0
2
0
0
0
0
0
1
1
0
0
1
NEGATIVE
NEUTRAL
POSITIVE
Predicted


Comparative Analysis of Model Performance Across Levels
Metrics
Accuracy
Precision
Recall
F1 Score
Score
1.0
0.8
0.6
0.4
0.2
0.0
Level 1
Level 2
Level 3
Level 4
Level

CONCLUSION

This project successfully evaluated and compared the performance of various classification models across different levels of a sentiment analysis task. Key findings include:

Model Performance: Different models exhibited varying effectiveness, with metrics such as accuracy, precision, recall, and F1-score providing a comprehensive assessment of their capabilities.

Comparative Insights: Visualizations highlighted performance differences across levels, guiding the selection of the most suitable model for specific tasks.

Future Directions: Opportunities for improvement include hyperparameter tuning, feature engineering, and exploring advanced models like transformers to enhance classification accuracy.

Real-World Applications: The insights gained can be applied in practical scenarios, such as sentiment analysis in social media and customer feedback, leading to better decision-making and user experiences.

THANK YOU!

As we look to the future, these technologies will play a crucial role in shaping a better world for generations to come.