WHAT IS CLAIMED IS:

1. A computer-implemented method, comprising:

performing one or more sampling operations on a training database that includes subjective scores to generate a plurality of resampled datasets;

for each resampled dataset, performing one or more machine learning operations based on the resampled dataset to generate a different bootstrap perceptual quality model; and

performing one or more operations based on the bootstrap perceptual models to quantify the accuracy of a perceptual quality score generated by a baseline perceptual quality model for a portion of encoded video content.

2. The computer-implemented method of claim 1, further comprising performing one or more machine learning operations based on the training database to generate the baseline perceptual quality model.

3. The computer-implemented method of claim 1, wherein each of the perceptual quality model and the bootstrap perceptual quality models maps a set of feature values for a set of features that characterize encoded video content to a different perceptual quality score.

4. The computer-implemented method of claim 3, wherein the set of features includes at least one of an additive impairment measure (ADM) feature, a visual information fidelity (VIF) feature, a blind/referenceless image spatial quality evaluator (BRISQUE) feature, and a temporal information (TI) feature.

5. The computer-implemented method of claim 1, wherein performing the one or more sampling operations comprises performing one or more resampling with replacement operations on feature values associated with portions of training encoded video content included in the training database.

6. The computer-implemented method of claim 1, wherein performing the one or more sampling operations comprises performing one or more resampling with replacement operations on the subjective scores to generate a plurality of resampled

subjective matrices.

7. The computer-implemented method of claim 1, wherein performing the one or more sampling operations comprises:

performing one or more resampling with replacement operations on feature values associated with portions of encoded content included in the training database to generate a plurality of resampled feature matrices; and

performing one or more resampling with replacement operations on the subjective scores to generate a plurality of resampled subjective matrices.

8. The computer-implemented method of claim 7, further comprising combining each resampled feature matrix included in the plurality of resampled feature matrices with each resampled subjective matrix included in the plurality of resampled subjective matrices to generate the plurality of resampled datasets.

9. The computer-implemented method of claim 1, wherein performing the or more operations based on the bootstrap perceptual models comprises, for each bootstrap perceptual model, generating a different bootstrap perceptual quality score based on the portion of encoded video content, wherein the distribution of the bootstrap perceptual quality scores quantifies the accuracy of the perceptual quality score.

10. The computer-implemented method of claim 1, wherein performing the or more operations based on the bootstrap perceptual models comprises:

for each bootstrap perceptual model, generating a different bootstrap perceptual quality score based on the portion of encoded video content; and

computing a confidence interval based on the bootstrap perceptual quality scores and a confidence level.

11. One or more non-transitory computer readable media including instructions that, when executed by one or more processors, cause the one or more processors to perform the steps of:

performing one or more sampling operations on a training database that includes subjective scores to generate a plurality of resampled datasets;

for each resampled dataset, performing one or more machine learning operations based on the resampled dataset to generate a different bootstrap perceptual quality model; and

computing a plurality of bootstrap perceptual quality scores based on the bootstrap perceptual models and a portion of encoded video content to quantify the accuracy of a perceptual quality score generated by a baseline perceptual quality model for the portion of encoded video content.

12 .The one or more non-transitory computer readable media of claim 11, wherein a first subjective score included in the subjective scores is assigned by a human in response to viewing a reconstructed version of a portion of encoded video content included in the training database.

13. The one or more non-transitory computer readable media of claim 11, wherein the perceptual quality score is a value for a Video Multimethod Assessment Fusion (VMAF) metric.

14. The one or more non-transitory computer readable media of claim 11, wherein performing the one or more sampling operations comprises performing one or more resampling with replacement operations on feature values associated with portions of training encoded video content included in the training database.

15. The one or more non-transitory computer readable media of claim 11, wherein performing the one or more sampling operations comprises performing one or more resampling with replacement operations on the subjective scores to generate a plurality of resampled subjective matrices.

16. The one or more non-transitory computer readable media of claim 15, further comprising performing one or more aggregation operations on each resampled subjective matrix included in the plurality of resampled subjective matrices to to generate a plurality of subjective vectors.

17. The one or more non-transitory computer readable media of claim 11, wherein performing the one or more sampling operations comprises:

performing one or more resampling with replacement operations on feature values associated with portions of encoded content included in the training database to generate a plurality of resampled feature matrices; and

performing one or more resampling with replacement operations on the subjective scores to generate a plurality of resampled subjective matrices.

18. The one or more non-transitory computer readable media of claim 17, further comprising combining each resampled feature matrix included in the plurality of resampled feature matrices with each resampled subjective matrix included in the plurality of resampled subjective matrices to generate the plurality of resampled datasets.

19. The one or more non-transitory computer readable media of claim 11, further comprising computing a confidence interval based on the plurality of bootstrap perceptual quality scores and a confidence level.

20. A system, comprising:

one or more memories storing instructions; and

one or more processors that are coupled to the one or more memories and, when executing the instructions, are configured to:

perform one or more resampling with replacement operations on a training database that includes subjective scores and feature values to generate a plurality of resampled datasets;

for each resampled dataset, perform one or more machine learning operations based on the resampled dataset to generate a different bootstrap perceptual quality model; and

perform one or more operations based on the bootstrap perceptual models to quantify the accuracy of a perceptual quality score generated by a baseline perceptual quality model for a portion of

encoded video content.

21. A computer-implemented method, comprising:

computing a first plurality of quality scores associated with a test encoding configuration based on a plurality of bootstrap quality models, wherein each bootstrap quality model is trained based on a different subset of a training database;

computing a second plurality of quality scores associated with a reference encoding configuration based on the plurality of bootstrap quality models; and

generating a distribution of bootstrap values for an encoding comparison metric based on the first plurality of quality scores and the second plurality of quality scores, wherein the distribution quantifies an accuracy of a baseline value for the encoding comparison metric generated by a baseline quality model.

22. The computer-implemented method of claim 21, wherein the encoding comparison metric comprises a Bjontegaard rate difference (BD-rate).

23. The computer-implemented method of claim 21, wherein each quality score included in the first plurality of quality scores is a different value for a Video Multimethod Assessment Fusion (VMAF) metric.

24. The computer-implemented method of claim 21, wherein the first encoding configuration is associated with a first coder/decoder (codec), and the second encoding configuration is associated with a second codec.

25. The computer-implemented method of claim 21, further comprising:

computing a third plurality of quality scores associated with the test encoding configuration based on the baseline quality model;

computing a fourth plurality of quality scores associated with reference configuration based on the baseline quality model; and

computing the baseline value for the encoding comparison metric based on the third plurality of quality scores and the fourth plurality of quality scores.

26. The computer-implemented method of claim 21, wherein computing the first plurality of quality scores comprises:

performing one or more encoding operations based on the test encoding configuration and a portion of video content, to generate a portion of encoded video content; and

computing a first quality score based on the portion of encoded video content and a first bootstrap quality model included in the plurality of bootstrap quality models.

27. The computer-implemented method of claim 21, wherein each quality score included in the first plurality of quality scores is associated with a different combination of a portion of encoded video content, a bitrate setting, and a bootstrap quality model included in the plurality of bootstrap quality models.

28. The computer-implemented method of claim 21, wherein the baseline quality model is trained based on the training database.

29. The computer-implemented method of claim 21, further comprising performing one or more content analysis operations based on the distribution of bootstrap values to quantify the performance of the test encoding configuration for a first type of video content.

30. The computer-implemented method of claim 21, further comprising performing one or more statistical analysis operations based on the distribution of bootstrap values and another distribution of bootstrap values associated with a different encoding configuration to compare the encoding performance of the test encoding configuration to the encoding performance of the different encoding configuration.

31. One or more non-transitory computer readable media including instructions that, when executed by one or more processors, cause the one or more processors to perform the steps of:

computing a first plurality of quality scores associated with a test encoding configuration based on a plurality of bootstrap quality models, wherein

each bootstrap quality model is trained based on a different subset of a training database;

computing a second plurality of quality scores associated with a reference encoding configuration based on the plurality of bootstrap quality models;

generating a baseline value for an encoding comparison metric based on a baseline quality model trained based on the training database, the test encoding configuration, and the reference encoding configuration; and

generating a distribution of bootstrap values for the encoding comparison metric based on the first plurality of quality scores and the second plurality of quality scores, wherein the distribution quantifies an accuracy of the baseline value for the encoding comparison metric.

32. The one or more non-transitory computer readable media of claim 31, wherein the encoding comparison metric specifies a percentage bitrate change when encoding using the test encoding configuration relative to encoding using the reference encoding configuration while maintaining the same quality score.

33. The one or more non-transitory computer readable media of claim 31, wherein each quality score included in the first plurality of quality scores is a different value for a Video Multimethod Assessment Fusion (VMAF) metric.

34. The one or more non-transitory computer readable media of claim 31, wherein the first encoding configuration is associated with a first coder/decoder (codec), and the second encoding configuration is associated with a second codec.

35. The one or more non-transitory computer readable media of claim 31, wherein computing the first plurality of quality scores comprises:

performing one or more encoding operations based on the test encoding configuration and a portion of video content, to generate a portion of encoded video content; and

computing a first quality score based on the portion of encoded video content and a first bootstrap quality model included in the plurality of bootstrap quality models.

36. The one or more non-transitory computer readable media of claim 31, wherein each quality score included in the first plurality of quality scores is associated with a different combination of a portion of encoded video content, a bitrate setting, and a bootstrap quality model included in the plurality of bootstrap quality models.

37. The one or more non-transitory computer readable media of claim 31, further comprising generating a confidence interval based on the distribution of bootstrap values and a confidence level.

38. The one or more non-transitory computer readable media of claim 31, further comprising performing one or more content analysis operations based on the distribution of bootstrap values to quantify the performance of the test encoding configuration for a first type of video content.

39. The one or more non-transitory computer readable media of claim 31, further comprising performing a Kolmogorov Smirnov test based on the distribution of bootstrap values and another distribution of bootstrap values associated with a different encoding configuration to compare the encoding performance of the test encoding configuration to the encoding performance of the different encoding configuration.

40. A system, comprising:

one or more memories storing instructions; and

one or more processors that are coupled to the one or more memories and, when executing the instructions, are configured to:

compute a first plurality of perceptual quality scores associated with a test codec based on a plurality of bootstrap quality models, wherein each bootstrap quality model is trained based on a different subset of a training database;

compute a second plurality of perceptual quality scores associated with a reference codec based on the plurality of bootstrap quality models; and

generate a distribution of bootstrap values for an encoding comparison

metric based on the first plurality of perceptual quality scores and the second plurality of perceptual quality scores, wherein the distribution quantifies an accuracy of a baseline value for the encoding comparison metric generated by a baseline quality model trained based on the training database.

WHAT IS CLAIMED IS:

1. A computer-implemented method, comprising:

receiving a plurality of training pairs and a subjective matrix that includes individual subjective scores provided by a plurality of subjective reviewers for the plurality of training pairs;

computing a first plurality of quality scores associated with a test encoding configuration based on a plurality of bootstrap quality models, wherein each bootstrap quality model included in the plurality of bootstrap quality models is trained based on at least one training pair included in the plurality of training pairs as well as a subset of the subjective matrix that includes individual subjective scores provided by a different subset of the plurality of subjective reviewers;

computing a second plurality of quality scores associated with a reference encoding configuration based on the plurality of bootstrap quality models; generating a distribution of bootstrap values for an encoding comparison metric based on the first plurality of quality scores and the second plurality of quality scores; and

determining an accuracy of a baseline value for the encoding comparison metric based on the distribution of bootstrap values, wherein the baseline value for the encoding comparison metric is generated by a baseline quality model that is trained based on the plurality of training pairs and the individual subjective scores provided by the plurality of subjective reviewers.

2. The computer-implemented method of claim 1, wherein the encoding comparison metric comprises a Bjontegaard rate difference (BD-rate).

3. The computer-implemented method of claim 1, wherein each quality score included in the first plurality of quality scores is a different value for a Video Multimethod Assessment Fusion (VMAF) metric.

4. The computer-implemented method of claim 1, wherein the test encoding configuration is associated with a first coder/decoder (codec), and the reference encoding configuration is associated with a second codec.

5. The computer-implemented method of claim 1, further comprising:

computing a third plurality of quality scores associated with the test encoding configuration based on the baseline quality model;

computing a fourth plurality of quality scores associated with the reference encoding configuration based on the baseline quality model; and

computing the baseline value for the encoding comparison metric based on the third plurality of quality scores and the fourth plurality of quality scores.

6. The computer-implemented method of claim 1, wherein computing the first plurality of quality scores comprises:

performing one or more encoding operations based on the test encoding configuration and a portion of video content, to generate a portion of encoded video content; and

computing a first quality score based on the portion of encoded video content and a first bootstrap quality model included in the plurality of bootstrap quality models.

7. The computer-implemented method of claim 1, wherein each quality score included in the first plurality of quality scores is associated with a different combination of a portion of encoded video content, a bitrate setting, and a bootstrap quality model included in the plurality of bootstrap quality models.

8. The computer-implemented method of claim 1, wherein the baseline quality model is trained based on the training database.

9. The computer-implemented method of claim 1, further comprising performing one or more content analysis operations based on the distribution of bootstrap values to quantify a performance of the test encoding configuration for a first type of video content.

10. The computer-implemented method of claim 1, further comprising performing one or more statistical analysis operations based on the distribution of bootstrap values and another distribution of bootstrap values associated with a different encoding configuration to compare an encoding performance of the test encoding configuration to an encoding performance of the different encoding configuration.

11. One or more non-transitory computer readable media including instructions that, when executed by one or more processors, cause the one or more processors to perform the steps of:

receiving a plurality of training pairs and a subjective matrix that includes individual subjective scores provided by a plurality of subjective reviewers for the plurality of training pairs;

computing a first plurality of quality scores associated with a test encoding configuration based on a plurality of bootstrap quality models, wherein each bootstrap quality model included in the plurality of bootstrap quality models is trained based on at least one training pair included in the plurality of training pairs as well as a subset of the subjective matrix that includes individual subjective scores provided by a different subset of the plurality of subjective reviewers;

computing a second plurality of quality scores associated with a reference encoding configuration based on the plurality of bootstrap quality models;

generating a baseline value for an encoding comparison metric based on a baseline quality model trained based on the plurality of training pairs and the individual subjective scores provided by the plurality of subjective reviewers, the test encoding configuration, and the reference encoding configuration;

generating a distribution of bootstrap values for the encoding comparison metric based on the first plurality of quality scores and the second plurality of quality scores; and

determining an accuracy of the baseline value for the encoding comparison metric based on the distribution of bootstrap values.

12. The one or more non-transitory computer readable media of claim 11, wherein the encoding comparison metric specifies a percentage bitrate change when encoding using the test encoding configuration relative to encoding using the reference encoding configuration while maintaining the same quality score.

13. The one or more non-transitory computer readable media of claim 11, wherein each quality score included in the first plurality of quality scores is a different value for a Video Multimethod Assessment Fusion (VMAF) metric.

14. The one or more non-transitory computer readable media of claim 11, wherein the test encoding configuration is associated with a first coder/decoder (codec), and the reference encoding configuration is associated with a second codec.

15. The one or more non-transitory computer readable media of claim 11, wherein computing the first plurality of quality scores comprises:

performing one or more encoding operations based on the test encoding configuration and a portion of video content, to generate a portion of encoded video content; and

computing a first quality score based on the portion of encoded video content and a first bootstrap quality model included in the plurality of bootstrap quality models.

16. The one or more non-transitory computer readable media of claim 11, wherein each quality score included in the first plurality of quality scores is associated with a different combination of a portion of encoded video content, a bitrate setting, and a bootstrap quality model included in the plurality of bootstrap quality models.

17. The one or more non-transitory computer readable media of claim 11, further comprising generating a confidence interval based on the distribution of bootstrap values and a confidence level.

18. The one or more non-transitory computer readable media of claim 11, further comprising performing one or more content analysis operations based on the distribution of bootstrap values to quantify a performance of the test encoding configuration for a first type of video content.

19. The one or more non-transitory computer readable media of claim 11, further comprising performing a Kolmogorov Smirnov test based on the distribution of bootstrap values and another distribution of bootstrap values associated with a different encoding configuration to compare an encoding performance of the test encoding configuration to an encoding performance of the different encoding configuration.

20. A system, comprising:

one or more memories storing instructions; and

one or more processors that are coupled to the one or more memories and, when executing the instructions, are configured to:

receive a plurality of training pairs and a subjective matrix that includes individual subjective scores provided by a plurality of subjective reviewers for the plurality of training pairs;

compute a first plurality of perceptual quality scores associated with a test codec based on a plurality of bootstrap quality models, wherein each bootstrap quality model included in the plurality of bootstrap quality models is trained based on at least one training pair included in the plurality of training pairs as well as a subset of the subjective matrix that includes individual subjective scores provided by a different subset of the plurality of subjective reviewers;

compute a second plurality of perceptual quality scores associated with a reference codec based on the plurality of bootstrap quality models;

generate a distribution of bootstrap values for an encoding comparison metric based on the first plurality of perceptual quality scores and the second plurality of perceptual quality scores; and

determine an accuracy of a baseline value for the encoding comparison metric based on the distribution of bootstrap values, wherein the baseline value for the encoding comparison metric is generated by a baseline quality model that is trained based on the plurality of training pairs and the individual subjective scores provided by the plurality of subjective reviewers.