



UNDERGRADUATE THESIS

MOBILEVIT-BASED DEEPPAKE DETECTION WITH SPATIAL-FREQUENCY FEATURE FUSION USING FFT AND MODIFIED CBAM

HABIB NURROHMAD SUGIHARTO
NPM 22081010165

THESIS ADVISORS

Dr. Faisal Muttaqin, S.Kom., M.T.
Budi Mukhamad Mulyo, S.Kom., M.T.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026**

APPROVAL SHEET

MOBILEVIT-BASED DEEPPAKE DETECTION WITH SPATIAL-FREQUENCY FEATURE FUSION USING FFT AND MODIFIED CBAM

By:
HABIB NURROHMAD SUGIHARTO
NPM. 22081010165

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on August 21, 2026:

Approved,

Dr. Faisal Muttaqin, S.Kom., M.T.
NIP. 19851231 202121 1 009


..... (Advisor I)

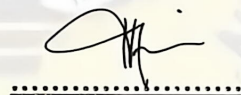
Budi Mukhamad Mulyo, S.Kom., M.T.
NIP. 19891118 202406 1 003


..... (Advisor II)

Dr. Eng. Ir. Anggraini Puspita Sari, S.T., M.T.
NPT. 222198 60 816400

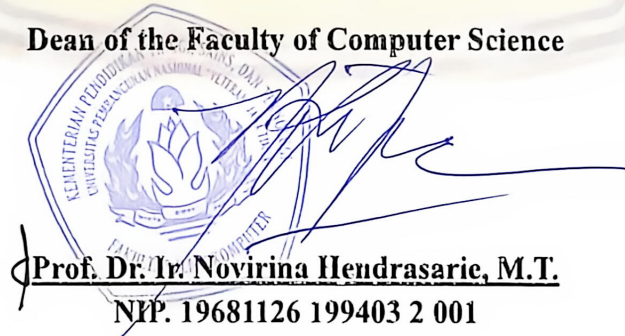

..... (Head Assessor)

Afina Lina Nurlaili, S.Kom., M.Kom.
NIP. 1993121 3202203 2 010


..... (Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science


Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 19681126 199403 2 001

APPROVAL SHEET

MOBILEVIT-BASED DEEPPAKE DETECTION WITH SPATIAL- FREQUENCY FEATURE FUSION USING FFT AND MODIFIED CBAM

By:
HABIB NURROHMAD SUGIHARTO
NPM. 22081010165

Approved to proceed to the Thesis Examination



Approved by,
Coordinator of Informatics Study Program
Faculty of Computer Science

Dr. Intan Yuniar Purbasari, S.Kom. MSc.
NIP. 19800602 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Habib Nurrohmah Sugiharto
NPM : 22081010165
Degree Program : Bachelor (S1)
Study Program : Informatics
Faculty : Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, September 14, 2026
Declarant,



HABIB NURROHMAD SUGIHARTO
NPM. 22081010165

ABSTRACT

Student Name / NPM: Habib Nurrohmad Sugiharto/22081010165
Thesis Title: MobileViT-Based Deepfake Detection With Spatial-Frequency Feature Fusion Using FFT and Modified CBAM
Supervisor: 1. Dr. Faisal Muttaqin, S.Kom., M.T.
2. Budi Mukhamad Mulyo, S.Kom., M.T.

The increasing realism of AI-generated deepfake content has made manipulated media more difficult to differentiate from genuine material. This challenge requires detection approaches that can maintain reliable classification performance, recognize subtle traces of manipulation, adapt to previously unseen data, and remain suitable for mobile-device implementation. This study proposes a deepfake detection model that combines MobileViT, Fast Fourier Transform (FFT), and Modified Convolutional Block Attention Module (CBAM). During preprocessing, YOLOv11n identifies the facial region and extracts it from each input frame. The extracted face is then represented through two complementary domains: RGB information provides the spatial representation, while FFT produces the frequency representation. These features are integrated using Fusion Convolution before further processing by MobileViT and Modified CBAM. FaceForensics++ (C23) serves as the secondary dataset, containing 20,000 frames equally divided into 10,000 real and 10,000 fake images. In addition, a primary dataset comprising 500 frames obtained from 50 videos is used to examine the model's ability to generalize beyond the training-data characteristics. Model performance is measured using accuracy, precision, recall, and F1-score. On the secondary dataset, the proposed MobileViT + FFT + Modified CBAM model achieves 95.30% accuracy, 0.9512 precision, 0.9550 recall, and an F1-score of 0.9531, exceeding the MobileViT baseline accuracy of 94.80%. Evaluation on the primary dataset yields 85.80% accuracy, 0.8594 precision, 0.8560 recall, and an F1-score of 0.8577 for the proposed model, whereas the baseline obtains 69.40% accuracy. These findings show that incorporating FFT and Modified CBAM into MobileViT strengthens classification performance, with a particularly substantial improvement when the model is evaluated on data with different characteristics. The proposed model is subsequently converted to TensorFlow Lite and deployed on-device within Telisik, a Flutter-based Android prototype developed for video deepfake detection.

Keywords: deepfake, MobileViT, FFT, Modified CBAM, Android

ACKNOWLEDGEMENTS

Praise be to Allah SWT for His grace and blessings, which have enabled the author to complete this thesis proposal titled **"MobileViT-Based Deepfake Detection With Spatial-Frequency Feature Fusion Using FFT and Modified CBAM"**. This research addresses the challenge of deepfake detection through spatial-frequency feature fusion. Ultimately, this proposal marks an important milestone in the author's academic journey toward contributing to computer vision research.

The preparation of this thesis proposal was made possible through the invaluable guidance, constructive feedback, and continuous support of numerous individuals. The author wishes to express sincere appreciation and gratitude to all parties who have devoted their time, knowledge, and encouragement throughout this process. In particular, the author would like to extend heartfelt thanks to:

1. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, M.T. as the Dean of the Faculty of Computer Science, National Development University "Veteran" East Java.
2. Mrs. Dr. Intan Yuniar Purbasari, S.Kom. M.Sc. as the Coordinator of Informatics Study Program, Faculty of Computer Science, National Development University of East Java.
3. Dr. Ir. Kartini, S.Kom., M.T., as the Academic Advisor, for her ongoing guidance, wise counsel, and encouragement throughout the author's academic journey.
4. Dr. Faisal Muttaqin, S.Kom., M.T., as Supervisor I, and Budi Mukhamad Mulyo, S.Kom., M.T., as Supervisor II, for their dedicated time, invaluable guidance, technical support, and continuous encouragement throughout the preparation of this research proposal.
5. Dr. Eng. Ir. Anggraini Puspita Sari, S.T., M.T., as the Head Assessor, and Afina Lina Nurlaili, S.Kom., M.Kom., as Assessor 1, for their valuable time, thorough evaluation, and constructive feedback in reviewing and perfecting this thesis proposal.
6. The Lecturers and Academic Staff of the Informatics Study Program and Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran"

Jawa Timur, for sharing their knowledge and guidance during the author's studies.

7. My dearest Mother, who has played an irreplaceable role in the author's life through her endless prayers, emotional warmth, and unconditional support. Her constant presence serves as an enduring source of strength and comfort whenever the author feels weary.
8. My dearest Father, who taught the author the fundamental values of responsibility, hard work, and resilience. His wise counsel, guidance, and personal examples serve as a solid foundation for navigating every challenge with determination.
9. Iyan, Sekar, Rencsa, and Zaldy, as close friends who have always been by the author's side, offering constant presence, unwavering support, and shared moments through every stage of this journey.
10. Nanda, Ade, Fajar, Rhimba, Cinta, and Dela, for their continuous encouragement, kindness, and assistance throughout the preparation of this proposal.
11. Wahyu, Basroil, Paskal, Kadek, Manda, Deva, and Bregas, for their mentorship, inspiring advice, and continuous motivation that kept the author moving forward.
12. All parties who cannot be mentioned individually, who have been part of the author's journey, provided support, and grown together in the academic environment.

The author realizes that this thesis proposal is far from perfect and still contains shortcomings. Therefore, constructive criticism, feedback, and suggestions from readers are highly welcomed for future improvement. Finally, the author hopes that this report will provide meaningful insights and be useful for all parties concerned.

Surabaya, September 14th 2026

Author

TABLE OF CONTENTS

APPROVAL SHEET	iii
APPROVAL SHEET	v
STATEMENT OF ORIGINALITY	vii
ABSTRACT	ix
ACKNOWLEDGEMENTS.....	xi
TABLE OF CONTENTS.....	xiii
LIST OF FIGURES	xix
LIST OF TABLES	xxvii
LIST OF CODE	xxix
CHAPTER I INTRODUCTION.....	1
1.1. Background	1
1.2. Research Questions	3
1.3. Research Objectives	3
1.4. Research Contributions	4
1.5. Research Limitations.....	4
CHAPTER II LITERATURE REVIEW	7
2.1. Previous Studies	7
2.2. Deepfake	15
2.3. Computer Vision	17
2.4. Deep Learning.....	18
2.5. Convolutional Neural Network.....	18
2.6. Transfer Learning.....	20
2.7. YOLOv11n for Face Detection.....	21
2.8. MobileViT.....	23

2.8.1.	Stem Convolution Block	24
2.8.2.	MobileNetV2 Block	24
2.8.3.	MobileViT Block	27
2.8.4.	Global Average Pooling	32
2.9.	RGB Image	33
2.10.	Fast Fourier Transform	34
2.10.1.	Grayscale Conversion	35
2.10.2.	Two-Dimensional DFT and FFT	35
2.10.3.	Magnitude Spectrum	36
2.10.4.	Frequency shifting.....	36
2.10.5.	Logarithmic Transformation	37
2.10.6.	FFT Spectrum Normalization	38
2.11.	Spatial and Frequency Feature Fusion	39
2.12.	Image Resizing	40
2.13.	Data Augmentation	42
2.13.1.	Random Rotation	42
2.13.2.	Horizontal Flip	43
2.13.3.	Color Jitter.....	44
2.14.	Attention Mechanism and CBAM	44
2.14.1.	Channel Attention	45
2.14.2.	Spatial Attention	47
2.14.3.	Convolutional Block Attention Module.....	48
2.14.4.	Modified CBAM	49
2.15.	Classification Head	50
2.16.	Model Performance Evaluation.....	51
2.16.1.	Accuracy	52

2.16.2.	Precision.....	52
2.16.3.	Recall.....	53
2.16.4.	F1-score.....	53
CHAPTER III METHODOLOGY		55
3.1.	Research Stages.....	55
3.2.	Dataset and Research Environment	56
3.2.1.	Research Equipment.....	56
3.2.2.	Dataset Collection	58
3.2.3.	YOLO Dataset Splitting.....	61
3.2.4.	FaceForensics++ Dataset Preparation	62
3.3.	Data Preprocessing.....	63
3.3.1.	YOLOv11n Model Training	63
3.3.2.	Video Frame Extraction from the Dataset	65
3.3.3.	Face Detection and Cropping.....	67
3.3.4.	Image Resizing.....	69
3.3.5.	Preprocessed Dataset Split	71
3.4.	Model Input Data Formation.....	72
3.4.1.	Data Augmentation	73
3.4.2.	RGB Image Normalization	81
3.4.3.	Frequency Image Formation	83
3.4.4.	Spatial-Frequency Feature Fusion	88
3.5.	Model Architecture Design.....	92
3.5.1.	MobileViT Backbone.....	92
3.5.2.	Modified CBAM Mechanism	106
3.5.3.	Pooling and Classification Layers.....	109
3.6.	Model Training and Validation Procedure.....	112

3.6.1.	Control Parameter Configuration and Model Training Scheme .	112
3.6.2.	Validation Strategy and Best Model Selection	114
3.7.	Testing Scenarios	115
3.8.	Android Implementation	116
3.8.1.	System Implementation Flow	117
3.8.2.	User Interface Design.....	119
CHAPTER IV RESULT AND DISCUSSION.....		123
4.1.	Data Preparation.....	123
4.1.1.	Secondary Dataset.....	123
4.1.2.	Primary Dataset.....	126
4.2.	Data Preprocessing Results	127
4.2.1.	YOLOv11n Model Training and Evaluation	127
4.2.2.	Video Frame Extraction	128
4.2.3.	Face Detection Using YOLOv11n.....	131
4.2.4.	Image Resizing.....	133
4.2.5.	Preprocessed Dataset Results and Split	134
4.3.	Model Input Data Formation.....	135
4.3.1.	Image Augmentation.....	135
4.3.2.	RGB Image Normalization	136
4.3.3.	Frequency Image Formation Using FFT.....	137
4.3.4.	Spatial-Frequency Feature Fusion	138
4.4.	Model Implementation	139
4.4.1.	MobileViT.....	139
4.4.2.	<i>Modified</i> CBAM	140
4.4.3.	Classification Head	142
4.5.	Hyperparameter Testing Scenarios	143

4.5.1.	Scenario 1: Testing the Preprocessed Data Split Proportion	144
4.5.2.	Scenario 2: Testing the Model Learning Rate.....	150
4.5.3.	Scenario 3: Testing the Number of Fusion Convolution Filters .	156
4.5.4.	Scenario 4: Testing the Initial α Value in Modified CBAM.....	162
4.5.5.	Scenario 5: Testing the Dropout Value.....	168
4.5.6.	Scenario 6: Testing the Number of Epochs.....	174
4.6.	Model Performance Evaluation.....	180
4.6.1.	Model Performance Comparison on Secondary Data.....	181
4.6.2.	Model Generalization Comparison on Primary Data.....	189
4.7.	Android Implementation	195
CHAPTER V CONCLUSION AND RECOMMENDATIONS.....		207
5.1.	Conclusion	207
5.2.	Recommendations	208
BIBLIOGRAPHY		209
APPENDIX.....		221

LIST OF FIGURES

Figure 2.1 Original Face and Deepfake	15
Figure 2.2 Visualization of the Deepfake Image Generation Process.....	16
Figure 2.3 Illustration of the Convolutional Neural Network Architecture	19
Figure 2.4 Illustration of the Transfer Learning Concept	20
Figure 2.5 MobileViT Architecture	23
Figure 2.6 MobileNetV2 Block in the MobileViT Architecture	25
Figure 2.7 MobileViT Block	28
Figure 2.8 Channel attention Model	46
Figure 2.9 Spatial attention Model.....	48
Figure 2.10 CBAM Workflow	48
Figure 2.11 Confusion Matrix for Binary Classification	51
Figure 3.1 Research Stage Framework	55
Figure 3.2 Dataset and Data Preparation Flow	56
Figure 3.3 Visualization of YOLOv11n Face Detection Dataset Samples	59
Figure 3.4 Visualization of the FaceForensics++ Dataset.....	60
Figure 3.5 Data Preprocessing Flow	63
Figure 3.6 Visualization of the RGB Input Image	67
Figure 3.7 9x9 Input Image Matrix for the Red Channel.....	68
Figure 3.8 9x9 Input Image Matrix for the Green Channel	68
Figure 3.9 9x9 Input Image Matrix for the <i>Blue</i> Channel.....	68
Figure 3.10 Source Pixel Area Contributing to <i>O1</i>	70
Figure 3.11 Resized Matrix (7×7 , Red channel)	70
Figure 3.12 Resized Matrix (7×7 , <i>Green</i> channel).....	71
Figure 3.13 Resized Matrix (7×7 , <i>Blue</i> channel)	71
Figure 3.14 Model Input Data Formation Flow	72
Figure 3.15 Previous Resized Pixel Matrix (7×7 , <i>Red channel</i>).....	73
Figure 3.16 Pixel Area Used in the Interpolation Process	74
Figure 3.17 Pixels after Rotation Augmentation (7×7 , <i>Red channel</i>).....	75
Figure 3.18 Pixels after Rotation Augmentation (7×7 , <i>Green</i> channel).....	75
Figure 3.19 Pixels after Rotation Augmentation (7×7 , <i>Blue</i> channel).....	76

Figure 3.20 Pixels after Horizontal Flip Augmentation (7×7 , <i>Red channel</i>)	77
Figure 3.21 Pixels after Horizontal Flip Augmentation (7×7 , <i>Green channel</i>).....	77
Figure 3.22 Pixels after Horizontal Flip Augmentation (7×7 , <i>Blue channel</i>)	77
Figure 3.23 Pixels after Brightness Augmentation (7×7 , <i>Red channel</i>).....	78
Figure 3.24 Pixels after Brightness Augmentation (7×7 , <i>Green channel</i>)	79
Figure 3.25 Pixels after Brightness Augmentation (7×7 , <i>Blue channel</i>).....	79
Figure 3.26 Pixels after Contrast Augmentation (7×7 , <i>Red channel</i>)	80
Figure 3.27 Pixels after Contrast Augmentation (7×7 , <i>Green channel</i>).....	80
Figure 3.28 Pixels after Contrast Augmentation (7×7 , <i>Blue channel</i>)	80
Figure 3.29 Normalized Matrix (7×7 , <i>Red channel</i>).....	82
Figure 3.30 Normalized Matrix (7×7 , <i>Green channel</i>)	82
Figure 3.31 Normalized Matrix (7×7 , <i>Blue channel</i>).....	83
Figure 3.32 7×7 Matrix after RGB-to-Grayscale Conversion	84
Figure 3.33 Two-Dimensional Fourier Transform and Magnitude Spectrum.....	85
Figure 3.34 Frequency Shifting Result	86
Figure 3.35 Logarithmic Transformation Result.....	87
Figure 3.36 Normalized FFT Result	88
Figure 3.37 Spatial-Frequency Feature Fusion Flow	88
Figure 3.38 Simulation of 3×3 Fusion Convolution Calculation.....	89
Figure 3.39 Feature Map Z after Fusion Convolution	90
Figure 3.40 Matrix Y after Projection Convolution	91
Figure 3.41 MobileViT + FFT + Modified CBAM Architecture.....	92
Figure 3.42 Input Matrix X from Projection Convolution with Padding 1.....	93
Figure 3.43 Element-Wise Multiplication of Input Patch X and Stem Convolution Kernel.....	94
Figure 3.44 Stem Convolution Output Matrix	94
Figure 3.45 MobileNetV2 Block Expansion Layer Output Matrix	95
Figure 3.46 Element-Wise Multiplication of F Expansion Input Patch and Depthwise Convolution Kernel.....	96
Figure 3.47 MobileNetV2 Block Depthwise Convolution Output Matrix	97
Figure 3.48 MobileNetV2 Block Projection Layer Output Matrix.....	97
Figure 3.49 MobileNetV2 Block Residual Connection Output Matrix	98

Figure 3.50 Element-Wise Multiplication of F MV2 Input Patch and Kernel in the MobileViT Block	99
Figure 3.51 MobileViT Local Convolution Output Matrix.....	99
Figure 3.52 Patches Generated by Unfolding the MobileViT Feature Map	100
Figure 3.53 MobileViT Global Feature Map after Folding.....	103
Figure 3.54 MobileViT Global Feature Map after Projection Layer	104
Figure 3.55 Element-Wise Multiplication of Initial Feature Patch F and Fusion Kernel.....	105
Figure 3.56 Element-Wise Multiplication of Projected Global Feature Patch G and Fusion Kernel.....	105
Figure 3.57 MobileViT Calculation Simulation Output Matrix.....	106
Figure 3.58 Modified CBAM Flow	107
Figure 3.59 Channel Attention Output Matrix in Modified CBAM	107
Figure 3.60 Spatial Attention Output Matrix in Modified CBAM	108
Figure 3.61 Modified CBAM Calculation Simulation Output Matrix.....	109
Figure 3.62 Telisik Application System Implementation Flow.....	117
Figure 3.63 Home Page.....	119
Figure 3.64 Video Range Selector Page.....	120
Figure 3.65 Detection Result Page.....	121
Figure 4.1 YOLOv11n Training Dataset on the Roboflow Website	124
Figure 4.2 FaceForensics++ Dataset on the Kaggle Website.....	124
Figure 4.3 FaceForensics++ Dataset Distribution Before and After Balancing .	125
Figure 4.4 Example of Facial Manipulation on Video Using NeoRefacer	126
Figure 4.5 YOLOv11n Model Validation Metrics	128
Figure 4.6 Frame Extraction Results on the Secondary Dataset.....	129
Figure 4.7 Frame Extraction Results on the Primary Dataset.....	130
Figure 4.8 Face Region Cropping Results on the Secondary Dataset	132
Figure 4.9 Face Region Cropping Results on the Primary Dataset	132
Figure 4.10 Image Resizing Results on the Secondary Dataset.....	133
Figure 4.11 Image Resizing Results on the Primary Dataset.....	133
Figure 4.12 Examples of Augmented Images	136
Figure 4.13 Example of an Image after FFT Transformation	138

Figure 4.14 Model Loss and Accuracy Curves for the 60:20:20 Ratio.....	144
Figure 4.15 Confusion Matrix Result for the 60:20:20 Ratio	145
Figure 4.16 Classification Report for the 60:20:20 Ratio	145
Figure 4.17 Model Loss and Accuracy Curves for the 70:15:15 Ratio.....	146
Figure 4.18 Confusion Matrix Result for the 70:15:15 Ratio	147
Figure 4.19 Classification Report for the 70:15:15 Ratio	147
Figure 4.20 Model Loss and Accuracy Curves for the 80:10:10 Ratio.....	148
Figure 4.21 Confusion Matrix Result for the 80:10:10 Ratio	148
Figure 4.22 Classification Report for the 80:10:10 Ratio	149
Figure 4.23 Model Loss and Accuracy Curves for Learning Rate 0.001.....	150
Figure 4.24 Confusion Matrix Result for Learning Rate 0.001	151
Figure 4.25 Classification Report for Learning Rate 0.001	151
Figure 4.26 Model Loss and Accuracy Curves for Learning Rate 0.0001.....	152
Figure 4.27 Confusion Matrix Result for Learning Rate 0.0001	152
Figure 4.28 Classification Report for Learning Rate 0.0001	153
Figure 4.29 Model Loss and Accuracy Curves for Learning Rate 0.00001.....	154
Figure 4.30 Confusion Matrix Result for Learning Rate 0.00001	154
Figure 4.31 Classification Report for Learning Rate 0.00001	155
Figure 4.32 Model Loss and Accuracy Curves for 8 Filters	156
Figure 4.33 Hasil <i>Confusion Matrix</i> Untuk 8 Filter.....	157
Figure 4.34 <i>Classification Report</i> Untuk 8 Filter	157
Figure 4.35 Model Loss and Accuracy Curves for 12 Filters	158
Figure 4.36 Confusion Matrix Result for 12 Filters.....	158
Figure 4.37 Classification Report for 12 Filters.....	159
Figure 4.38 Model Loss and Accuracy Curves for 16 Filters	160
Figure 4.39 Confusion Matrix Result for 16 Filters.....	160
Figure 4.40 Classification Report for 16 Filters.....	161
Figure 4.41 Model Loss and Accuracy Curves for Alpha 0.....	162
Figure 4.42 Confusion Matrix Result for Alpha 0	163
Figure 4.43 Classification Report for Alpha 0	163
Figure 4.44 Model Loss and Accuracy Curves for Alpha 0.5	164
Figure 4.45 Confusion Matrix Result for Alpha 0.5	164

Figure 4.46 Classification Report for Alpha 0.5	165
Figure 4.47 Model Loss and Accuracy Curves for Alpha 1	166
Figure 4.48 Confusion Matrix Result for Alpha 1	166
Figure 4.49 Classification Report for Alpha 1	167
Figure 4.50 Model Loss and Accuracy Curves for Dropout 0.3	168
Figure 4.51 Confusion Matrix Result for Dropout 0.3	169
Figure 4.52 Classification Report for Dropout 0.3	169
Figure 4.53 Model Loss and Accuracy Curves for Dropout 0.4	170
Figure 4.54 Confusion Matrix Result for Dropout 0.4	170
Figure 4.55 Classification Report for Dropout 0.4	171
Figure 4.56 Model Loss and Accuracy Curves for Dropout 0.5	172
Figure 4.57 Confusion Matrix Result for Dropout 0.5	172
Figure 4.58 Classification Report for Dropout 0.5	173
Figure 4.59 Model Loss and Accuracy Curves for 30 Epochs.....	174
Figure 4.60 Confusion Matrix Result for 30 Epochs	175
Figure 4.61 Classification Report for 30 Epochs	175
Figure 4.62 Model Loss and Accuracy Curves for 40 Epochs.....	176
Figure 4.63 Confusion Matrix Result for 40 Epochs	176
Figure 4.64 Classification Report for 40 Epochs	177
Figure 4.65 Model Loss and Accuracy Curves for 50 Epochs.....	178
Figure 4.66 Confusion Matrix Result for 50 Epochs	178
Figure 4.67 Classification Report for 50 Epochs	179
Figure 4.68 Loss and Accuracy Curves of the MobileViT (Baseline) Model on Secondary Data	181
Figure 4.69 Confusion Matrix of the MobileViT (<i>Baseline</i>) Model on Secondary Data	182
Figure 4.70 Classification Report of the MobileViT (Baseline) Model on Secondary Data	182
Figure 4.71 Loss and Accuracy Curves of the MobileViT + FFT Model on Secondary Data	183
Figure 4.72 Confusion Matrix of the MobileViT + FFT Model on Secondary Data	183

Figure 4.73 Classification Report of the MobileViT + FFT Model on Secondary Data	184
Figure 4.74 Loss and Accuracy Curves of the MobileViT + Modified CBAM Model on Secondary Data.....	185
Figure 4.75 Confusion Matrix of the MobileViT + Modified CBAM Model on Secondary Data	185
Figure 4.76 Classification Report of the MobileViT + Modified CBAM Model on Secondary Data	186
Figure 4.77 Loss and Accuracy Curves of the MobileViT + FFT + Modified CBAM (Proposed) Model on Secondary Data	187
Figure 4.78 Confusion Matrix of the MobileViT + FFT + Modified CBAM (Proposed) Model on Secondary Data	187
Figure 4.79 Classification Report of the MobileViT + FFT + Modified CBAM (Proposed) Model on Secondary Data	188
Figure 4.80 Confusion Matrix of the MobileViT (Baseline) Model on Primary Data	190
Figure 4.81 Classification Report of the MobileViT (Baseline) Model on Primary Data	190
Figure 4.82 Confusion Matrix of the MobileViT + FFT Model on Primary Data	191
Figure 4.83 Classification Report of the MobileViT + FFT Model on Primary Data	191
Figure 4.84 Confusion Matrix of the MobileViT + Modified CBAM Model on Primary Data	192
Figure 4.85 Classification Report of the MobileViT + Modified CBAM Model on Primary Data	193
Figure 4.86 Confusion Matrix of the MobileViT + FFT + Modified CBAM (Proposed) Model on Primary Data	193
Figure 4.87 Classification Report of the MobileViT + FFT + Modified CBAM (Proposed) Model on Primary Data	194
Figure 4.88 Home Page.....	200
Figure 4.89 Video Range Selector Page.....	201

Figure 4.90 Result Page (Deepfake Video).....	202
Figure 4.91 Result Page (Original Video).....	203
Figure 4.92 Video Detection Results with Animal Objects	204
Figure 4.93 Video Detection Results with Non-Living Objects	205

LIST OF TABLES

Table 2.1 Previous Studies	11
Table 3.1 Research Laptop Specifications	57
Table 3.2 Research Smartphone Specifications	57
Table 3.3 Experimental Environment and Software	58
Table 3.4 Summary of Datasets Used	60
Table 3.5 YOLO Model Training Dataset Distribution.....	61
Table 3.6 FaceForensics++ Dataset Distribution for Preprocessing	62
Table 3.7 YOLO Model Training Hyperparameters	64
Table 3.8 Estimated Frame Extraction Results from the FaceForensics++ Dataset	66
Table 3.9 Estimated Frame Extraction Results from the Primary Dataset.....	66
Table 3.10 Distribution of the Preprocessed FaceForensics++ Dataset under Different Split Ratios	72
Table 3.11 Training Control Parameter Configuration.....	112
Table 3.12 Parameters Evaluated in the Testing Scenarios	113
Table 3.13 Model Parameter Testing Scenarios	115
Table 4.1 Dataset Split for YOLOv11n Model Training	124
Table 4.2 Data Equalization Results for the FaceForensics++ Dataset	125
Table 4.3 Primary Dataset Collection Results.....	126
Table 4.4 Frame Extraction Results on the Secondary Dataset	130
Table 4.5 Frame Extraction Results on the Primary Dataset	131
Table 4.6 Preprocessed Dataset Split Results	134
Table 4.7 Composition of the Preprocessed Primary Dataset	135
Table 4.8 Comparison of Data Split Proportion Testing Results	149
Table 4.9 Comparison of Learning Rate Testing Results.....	155
Table 4.10 Comparison of Fusion Convolution Filter Testing Results	161
Table 4.11 Comparison of Initial α Value Testing Results	167
Table 4.12 Comparison of Dropout Value Testing Results	173
Table 4.13 Comparison of Epoch Testing Results	179
Table 4.14 Optimal Model Parameter Configuration.....	180

Table 4.15 Comparison of Model Metrics on Secondary Data	188
Table 4.16 Comparison of Model Metrics on Primary Data	194

LIST OF CODE

Program Code 4.1 YOLOv11n Model Training.....	127
Program Code 4.2 Determining Frame Indices Using Uniform Frame Sampling	128
Program Code 4.3 Video Frame Extraction Process	129
Program Code 4.4 Face Detection Implementation Using YOLO.....	131
Program Code 4.5 Face Cropping Process Using YOLO Detection Results.....	132
Program Code 4.6 Image Resizing Implementation	133
Program Code 4.7 Data Augmentation Implementation	136
Program Code 4.8 RGB Normalization	137
Program Code 4.9 FFT Transformation	137
Program Code 4.10 RGB and FFT Fusion.....	138
Program Code 4.11 MobileViT Implementation.....	139
Program Code 4.12 Channel Attention Module.....	140
Program Code 4.13 Spatial Attention Module	141
Program Code 4.14 Modified CBAM Integration	142
Program Code 4.15 Classification Head Implementation.....	143
Program Code 4.16 Sigmoid Activation Function	143
Program Code 4.17 Conversion of the .pt Model to .tflite.....	196
Program Code 4.18 Flutter Application Initialization.....	196
Program Code 4.19 Page Navigation Structure	197
Program Code 4.20 Video Input Interface and Scan Trigger	197
Program Code 4.21 Scanning Pipeline: Frame Sampling and Face Selection....	198
Program Code 4.22 Scanning Pipeline: Data Transfer to the Result Page	198
Program Code 4.23 Batch Inference in the Result Controller	199
Program Code 4.24 Probability Aggregation and Video Classification Decision	199