Ministry of Higher
Education and Scientific Research
University of Diyala
College of science
Department of Computer Science

DeepFake Detection System

**A Thesis Submitted to the Department of Computer Science/ College of Science/ University of Diyala
In Partial Fulfilment of the Requirements for the Degree of MASTER in Computer Science**

By

Mohammed Akram Younus Al-Sa'ati

Supervised By

Asst. Prof. Dr .Taha Mohammad Hasan

2020AC **1442AH**

بسم الله الرحمن الرحيم

﴿ نَرْفَعُ دَرَجَاتٍ مَّن نَّشَاءُ وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ ﴾

صدق الله العظيم

سورة يوسف

الاية (٧٦)

Acknowledgments

First of all, praise is to ***Allah*** *the lord of the whole creation, on all the blessing was the help in achieving this research to its end.*

I wish to express my thanks to my supervisors ***Dr. Taha Mohammad Hassan*** *for supervising this research and for the generosity, patience, and continuous guidance throughout the work. It has been my good fortune to have the advice and guidance from him. My thanks to the academic and administrative staff at the Department of the computer sciences\University of Diyala for their hospitality and generosity.*

I would like to express my gratitude to my ***wife*** *and* ***daughters*** *for their great support and continuous encouragement. Not to forget to mention the great generosity of* ***Mr. Khalil Ibrahim Abed Al-Qaisi****. This accomplishment would not have been possible without them.*

Dedication

To...

My father soul Prof. Dr.Akram Al-saati

My dear mother

My dear wife & my daughters

My Dear friend Mr.Khalil Al-Qaisi

All our distinguished teachers those who paved the way for our science and knowledge

Mohammed Akram Younus

(Supervisor's Certification)

We certify that this research entitled "DeepFake Detection System" was prepared by **Mohammed Akram Younus** under our supervision at the University of Diyala Faculty of Science Department of Computer Science, as partial fulfillment of the requirement needed to award the degree of Master of Science in Computer Science.

Signature: 

Name: Asst. Prof. Dr.Taha Mohammad Hasan

Date:

Approved by the University of Diyala Faculty of Science Department of Computer Science.

Signature: 

Name: Asst. Prof. Dr.Taha Mohammad Hasan

Date:

(Head of Computer Science Department)

(Linguistic Certification)

I certify that this research entitled “DeepFake Detection System” was prepared by Mohammed Akram Younus and was reviewed linguistically. Its language was amended to meet the style of the English language.

Signature:

Name:

Date:

Abstract

The easy and free entrance to large-scale public databases with the rapid quick progress of deep learning techniques and applying artificial intelligence (AI), especially the Generative Adversarial Networks (GAN), have driven to the production of very realistic fake videos content and guarantee an advanced level of realism with its implications towards corresponding society. This unlocks the door to a chain of sensational applications in different fields such as video games, film production, and advertising. On the other hand, it constitutes enormous security threats. Freely available software packages on the web allow any person, with minimum skills, to produce very realistic DeepFakes videos. This technology is used to blackmail and discredit people, manipulate the opinion of the public during elections, etc. There are no limits to the potential abuses of the human imagination. Subsequently, there is an imperious need for automated tools that have the capability of detecting such fake videos and averting the diffusion of dangerous fake multimedia content.

Previous works and researches in that field showed a lot of complexity with different accuracy. In this thesis, a new technique is described, proposing the use of the Wavelet Transform to determine the blur extent of the face region of interest (RIO) and the surrounding context by the use of edges type and make a comparison between them. This approach can successfully distinguish AI-generated counterfeit videos from genuine ones.

Based on the observations that the prevailing DeepFake set of rules can only generate images of constrained resolutions for the synthesized faces, which require additional distortion and obscuration to coordinate the valid appearances in the source video. Certainly, such changes will create unmistakable artifacts, the proposed method can effectively capture these artifacts by revealing the edge types and blurriness ratio. Most of the previous approaches need a large amount of DeepFake generated images and real portrayal to train the convolutional neural network (CNN). This technique does not need to bother with instances of

DeepFake produced depiction as negative preparation since it focuses on the artifacts as a distinctive feature in affine face warping to recognize if the videos are real or fake. Thus, economizes resource-demanding and time-consuming.

The used technique is more robust compared to others where such artifacts are general existed in DeepFake videos. The proposed method in this thesis was conducted using on the UADFV dataset and the result of the whole experiment result gave very good accuracy with great reliability, reached 100% on the mentioned dataset with a few nuances in each video test.

TABLE OF CONTENTS

	Contents	Page No.
	Abstract	I
	Table of contents	III
	List of Tables	V
	List of figures	V
	List of abbreviation	IX
	Chapter One: General Introduction	1-10
1.1	Introduction	1
1.2	DeepFake videos	2
1.3	Related work	3
1.4	Statement of the problem	9
1.5	The aim of the thises	9
1.6	Thesis organization	10
	Chapter Two: Theoretical Background	11-33
2.1	Introduction	11
2.2	History of DeepFake	12
2.3	The technological system of DeepFakes	13
2.4	Image Forensics	14
2.5	Computer Vision	15
2.6	Generative Adversarial Networks (GANs)	16
2.7	The Convolutional Neural Networks	18
2.8	Visual Artifacts	19
2.9	ResNet-50 for classification the image	20
2.10	Architecture of ResNet-50	20
2.11	Discrete Wavelets Transform	21
2.11.1	Edge detection problems	23
2.11.2	Using Haar wavelet for edge detection	24
2.11.3	HWT for edge detection Algorithm	27
2.12	Face detection	28
2.13	Histogram of Directional Gradients (HOG)	30
2.14	TensorFlow Hub & Transfer Learning	31
2.15	Results normalization	32
2.15.1	The norm of a vector	32
2.15.2	Normalization with the norm	33

	Chapter Three: The Proposed System	34-49
3.1	Introduction	34
3.2	The proposed system	34
3.2.1	Video pre-processing stage	36
3.2.2	Face Detection Stage	39
3.2.3	Blur Extent Detection Stage	41
3.2.4	Robustness Stage	45
3.2.4.B	ResNet-50 configuration	46
3.2.4.B	Using the ResNet-50 for Detecting Artifacts	47
3.2.5	Detecting Fake Video	48
	Chapter Four: The Experimental Test and Results	50-82
4.1	Introduction	50
4.2	UADFV dataset	50
4.3	Experimentation results	50
4.4	Results Evaluation	53
4.5	Test sample	82
	Chapter Five: Conclusions and Future Works	83-85
5.1	Conclusion	83
5.2	Future Work	84
	References	86-88

List of Tables

Table No.	Caption	Page No.
1.1	Summary of well-known DeepFake applications	3
2.1	Effect of HWT on different types of edges	27
3.1	ResNet-50 configuration	46
4.1	The parameters result in real and fake videos of the group (1)	53
4.2	The parameters result in real and fake videos of the group (2)	56
4.3	A comparison of DeepFake methods that are tested with the UADFV dataset	82

List of Figures

Table No.	Caption	Page No.
2.1	Lincoln-Calhoun Composite – Iconic Photo	12
2.2	Block diagram of the Generative Adversarial Network	17
2.3	A simple ResNet-50 residual block compared to standard network	21
2.4	Discrete wavelets transform tree.	22
2.5	Two-dimensional pictures Wavelet decomposition	22
2.6	One-level-of-DWT-decomposition-for-Akiyo-video-sequence_Q320.	25
2.7	Graphic description of edges type	26
2.8	Blur detection Structure scheme	26
2.9	Reference points for recognizing the face zone	29
2.10	A simple processed by Dlib's shape predictor	29
2.11	Example of the HOG-structure of the face	30
2.12	Transfer learning module.	32
3.1	General block diagram illustrates the system workflow	35
3.2	ResNet-50 architecture	46
3.3	A block diagram representation of pre-trained Resnet-50 architecture	46
4.1	The difference between the real and fake patterns of blur extent in real and fake cases. On the left, the results of the real video 0000.mp4. On the right, the fake video 0000_fake.mp4.	52
4.2	The face blur extent & context blur extent, probability of being fake, and blur differences for video "0000.mp4". (a) Real video-version, and (b) Fake video-version.	57
4.3	The face blur extent & context blur extent, probability of being fake, and blur differences for video "0001.mp4". (a) Real video-version, and (b) Fake video-version.	58
4.4	The face blur extent & context blur extent, probability of being fake, and blur differences for video "0002.mp4". (a) Real video-version, and (b) Fake video-version.	58
4.5	The face blur extent & context blur extent, probability of being fake, and blur differences for video "0003.mp4". (a) Real video-version, and (b) Fake video-version.	59

4.6	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0004.mp4”. (a) Real video-version, and (b) Fake video-version.	59
4.7	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0005.mp4”. (a) Real video-version, and (b) Fake video-version.	60
4.8	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0006.mp4”. (a) Real video-version, and (b) Fake video-version.	60
4.9	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0007.mp4”. (a) Real video-version, and (b) Fake video-version.	61
4.10	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0008.mp4”. (a) Real video-version, and (b) Fake video-version.	61
4.11	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0009.mp4”. (a) Real video-version, and (b) Fake video-version.	62
4.12	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0010.mp4”. (a) Real video-version, and (b) Fake video-version.	62
4.13	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0011.mp4”. (a) Real video-version, and (b) Fake video-version.	63
4.14	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0012.mp4”. (a) Real video-version, and (b) Fake video-version.	63
4.15	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0013.mp4”. (a) Real video-version, and (b) Fake video-version.	64
4.16	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0014.mp4”. (a) Real video-version, and (b) Fake video-version.	64
4.17	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0015.mp4”. (a) Real video-version, and (b) Fake video-version.	65
4.18	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0016.mp4”. (a) Real video-version, and (b) Fake video-version.	65
4.19	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0017.mp4”. (a) Real video-version, and (b) Fake video-version.	66
4.20	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0018.mp4”. (a) Real video-version, and (b) Fake video-version.	66
4.21	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0019.mp4”. (a) Real video-version, and (b) Fake video-version.	67

4.22	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0020.mp4”. (a) Real video-version, and (b) Fake video-version.	67
4.23	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0021.mp4”. (a) Real video-version, and (b) Fake video-version.	68
4.24	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0022.mp4”. (a) Real video-version, and (b) Fake video-version.	68
4.25	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0023.mp4”. (a) Real video-version, and (b) Fake video-version.	69
4.26	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0024.mp4”. (a) Real video-version, and (b) Fake video-version.	69
4.27	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0025.mp4”. (a) Real video-version, and (b) Fake video-version.	70
4.28	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0026.mp4”. (a) Real video-version, and (b) Fake video-version.	70
4.29	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0027.mp4”. (a) Real video-version, and (b) Fake video-version.	71
2.30	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0028.mp4”. (a) Real video-version, and (b) Fake video-version.	71
4.31	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0029.mp4”. (a) Real video-version, and (b) Fake video-version.	72
4.32	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0030.mp4”. (a) Real video-version, and (b) Fake video-version.	72
4.32	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0031.mp4”. (a) Real video-version, and (b) Fake video-version.	73
4.33	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0032.mp4”. (a) Real video-version, and (b) Fake video-version.	73
4.35	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0033.mp4”. (a) Real video-version, and (b) Fake video-version.	74
4.36	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0034.mp4”. (a) Real video-version, and (b) Fake video-version.	74
4.37	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0035.mp4”. (a) Real video-version, and (b) Fake video-version.	75

4.38	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0036.mp4”. (a) Real video-version, and (b) Fake video-version.	75
4.39	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0037.mp4”. (a) Real video-version, and (b) Fake video-version.	76
4.40	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0038.mp4”. (a) Real video-version, and (b) Fake video-version.	76
4.41	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0039.mp4”. (a) Real video-version, and (b) Fake video-version.	77
4.42	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0040.mp4”. (a) Real video-version, and (b) Fake video-version.	77
4.43	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0041.mp4”. (a) Real video-version, and (b) Fake video-version.	78
4.44	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0042.mp4”. (a) Real video-version, and (b) Fake video-version.	78
4.45	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0043.mp4”. (a) Real video-version, and (b) Fake video-version.	79
4.46	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0044.mp4”. (a) Real video-version, and (b) Fake video-version.	79
4.47	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0045.mp4”. (a) Real video-version, and (b) Fake video-version.	80
4.48	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0046.mp4”. (a) Real video-version, and (b) Fake video-version.	80
4.49	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0047.mp4”. (a) Real video-version, and (b) Fake video-version.	81
4.50	The face blur extent & context blur extent, probability of being fake, and blur differences for video “0048.mp4”. (a) Real video-version, and (b) Fake video-version.	81
4.51	Sample test of the proposed method on one of the “UADFV” DeepFake dataset.	82

List of Abbreviations

Abbreviations	Meaning
AI	Artificial Intelligence
ACC	Accuracy
API	Application Programming Interface
AUC	Area Under Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
CNN	Convolutional Neural Network
CPU	Central Processing Unit
Dlib	Digital Library
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
IU	Image Understanding
LRCN	Long-term Recurrent Convolutional Network
LSTM	Long Short Term Memory
MCP	McCulloch and Pitts
MTCNN	Multi-Task Cascade Networking
PRNU	Photo Response Non-Uniformity
ResNet	Residential Network
ROI	Region of Interest
SVM	Support-Vector Machines
VGG	Visual Geometry Group
HOG	Histogram of Oriented Gradients

Chapter One

General Introduction

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Introduction

In the last few years, the issue of fake news has turned into a threat to crowd discourse, human society, and democracy [1] [2]. Counterfeit news is a type of imaginative news-style to deceive the public with content that is fabricated [3] [4]. In 2020, 4.57 billion people were active internet users, encompassing 59 percent of the global population [5] [6], with a very large amount of images and videos are uploaded to the Internet each day. This includes millions of photos and over 400 hours of video content uploaded to social media and YouTube every minute. False news diffuses very fast especially via social media platforms, it can influence a very large number of internet users [7] where the studies showed that over 20% of the users follow the news through YouTube and Facebook [8].

The advanced technology of visual media advances has led to new facilities for processing and generating artificial videos. In particular, modern AI-based tools have been provided to create extremely hyper-realistic manipulated videos which are named DeepFakes, this new technology may comprise a serious threat to attack the general opinion on a certain event or the reputation of some individuals as almost with an average computer specification anyone can fabricate fake videos so easily that are virtually indistinguishable from actual real media [9]. Due to the rise in popularity of the misinformation the need to individuate this type of fake information becomes fundamental, a tool to confirm media news content genuinely, as new technologies permit persuasive manipulation of video [8]. Nowadays, the public lives in the era of what some have called a post-truth, it is characterized as information warfare running false news campaigns to manipulate public opinion driven by pernicious actors [10].

The proposed method shows a new forensic technique that can distinguish between fake and real video sequences with good accuracy. In this work the adoption of the inconsistency of the blur feature in the video frame to exploit the possible difference between ROI and the surrounded context. This clue is then used as a feature to be boosted by ResNet-50 classifiers. Preliminary results obtained on the DeepFake video dataset UADFV [11] [12] highlights very promising performances.

1.2 DeepFake videos

The term DeepFake is a mixture of Deep learning and Fake, DeepFakes are digitally manipulated hyper-realistic videos to depict individuals doing and saying things that have not occurred in reality. The production of DeepFakes depends mainly on artificial intelligence neural networks which are known as Generative Adversarial Networks (GANs) [13]. This neural network can learn in a comparable way to the brain of human beings. The more photos and videos of a person exposed to a neural network, the more accurately it can replicate the expressions of facial, and mannerisms of that person when producing a DeepFake video.

The most famous DeepFake applications and their highlights are introduced in Table (1.1). The procedure of DeepFake demand feeding two images (source and target) of different people into a Deep Learning [14] algorithm and train it to alternate their faces. It uses Artificial Intelligence and mapping of facial technology that swaps the face of a source person on a video into the face of the target person in the same video.

Table 1.1: Summary of well-known DeepFake applications.

Application	Key features
Faceswap	Shared parameters of the encoder are used. Uses two encoder-decoder pairs.
Faceswap-GAN	Perceptual misfortune using Visual Geometry Group (VGG) and adversarial disposed misfortune are added to the auto-encoder design.
DeepFaceLab	Bolster different face extraction modes, for example, Digital Library (Dlib), Multi-task Cascaded Convolutional Neural Networks (MTCNN), and Single Shot Scale-invariant Face Detector (S3FD). Grow the Faceswap model with new models.
DFaker	Implemented based on Keras library. DSSIM misfortune work is utilized to recreate the face.
DeepFake-TensorFlow	Same as DFaker but implemented depending on TensorFlow.

1.3 Related work

The detection of manipulated videos is much harder than fake image detection due to the solid corruption and degradation of the data of the frame after video compression [15]. A great challenge for methods designed to detect fake videos because of its temporal attributes that are differed among frames sets. Several related methods to the proposed work in this study have been reviewed in the literature.

- **P. Zhou et al. [16], 2017,** proposed a two-stream network for face tampering detection. They train GoogLeNet to detect tampering artifacts in a face classification stream and train a patch-based triplet network to leverage features capturing local noise residuals and camera characteristics as a second stream. Also, they used two different online face-swapping applications to create a new dataset that consists of 2010 tampered images, each of which contains a tampered face. They evaluated the proposed two-stream network on the newly collected dataset. Experimental results demonstrate the effectiveness of their method and it reaches 85.1% AUC. The Two-stream network is considered as a complex method and hard to train compared to the

results obtained. The proposed method was tested on the Celeb-DF and the results were too low, 53.8% AUC.

- **D. Afchar, et al. [15], 2018,** introduced a method to detect facial tampering in videos automatically and effectively at the mesoscopic level and focuses in particular on two recent techniques used to produce hyper-realistic faked videos, DeepFake and Face2Face. Due to the compression which severely degrades the data, traditional image forensics techniques are typically not well suited to videos. Thus, this work follows a deep learning approach at the mesoscopic level and presents two networks, both of which have a small number of layers to focus on image mesoscopic properties. A modified version of "Meso-4" is composed of a derivative of the "Inception module" proposed in [17], called "MesoInception-4". The solution they suggested was evaluated with a private dataset, achieving 98% ACC for the best accuracy results. The method was calibrated against unseen datasets in [18] and in some cases, as with "FaceForensics++", proved to be a robust approach, but its weakness was in finding the artifacts in some Deepfake videos, as seen in the results of the UADFV dataset which was 84.3% AUC.
- **M. Koopman, et al. [19], 2018,** proposed a method that uses the analysis of Photo Response Non-Uniformity (PRNU) to detect DeepFake video. The PRNU is characterized as an industrial facility deformity of light-sensitive sensors of advanced cameras known as noise patterns stemming. Every digital camera has its PRNU patterns and is considered as the digital image fingerprint [20] [21]. The exchanged faces are assumed to change the native PRNU patterns of video frames in the facial zone. The process starts by decomposing the videos into frames, then the facial regions are cropped. After that, it is then isolated sequentially into eight groups and an average of PRNU patterns is calculated for everyone. The results indicate that there is no correlation between the authenticity of the video and the variance in correlation scores. There does appear to be a correlation between the mean

correlation scores and the authenticity of the video, where on average original videos have higher mean normalized cross-correlation scores compared to the DeepFakes. Their proposed approach was evaluated over a private database created using 5 different mobile applications, achieving an average of 13.7% EER in manipulation detection which is considered a relatively high ratio compared to other methods.

- **M. Chang, et al. [22], 2018,** Proposed the Eye Blinking method to detect DeepFakes. Because of the way that an individual in DeepFakes has no regular eye blinking like that in non-manipulated videos. Normally images available on the internet do not show individuals with shut eyes, without having such images, DeepFake calculations cannot create scenes with faces that have normally blinking eyes. Distinguishing original from tampered videos, the author extracted the frames from the videos after that the eye areas are separated depending on six eye landmarks from face areas. Long-term recurrent convolutional network (LRCN) [23] is used on the cropped eye area sequences for the prediction of dynamic state. Based on CNN, the LRCN consists of a feature extractor, the arrangement learning relies upon Long Short Term Memory (LSTM), and the state forecast depends on the completely associated layer to gauge the probability of a shut-eye and open-eye state. Strong temporal dependencies were shown by the eye blinking, the LSTM implementation helps to capture these temporal patterns in a very effective way. The methodology was appraised on a lot of information that was gathered from the web comprising of 49 meetings and introduction recordings and their indistinguishable phony recordings delivered by the DeepFake algorithms. A promising result was gained from the proposed approach in detecting DeepFake videos, also, improvement can be performed by thinking about the dynamic pattern of blinking where exceedingly common blinking of the eyes might be considered as a sign of tampering. No long time

after this forensic technique was announced to the public, the upcoming age of synthesis strategies consolidated blinking into their frameworks.

- **Y. Li and S. J. Lyu [24], 2018,** proposed an approach that is based on the observations that the current DeepFake algorithm could only generate images of restricted resolutions, that need to be further warped to recreate the actual faces in the source video. These transforms leave distinctive artifacts in the resulting DeepFake videos, and they can be effectively captured by convolutional neural networks (CNNs). This approach was assessed on two DeepFake video datasets, called the DeepFakeTIMIT [25] and UADFV [12] [11]. The DeepFakeTIMIT dataset includes a set of 64×64 size low-resolution quality videos of and a second set of 128×128 high- resolution quality videos of with an around 10537 original images and 34,023 faked images obtained from 320 videos for each set of quality. The UADFV dataset consists of 49 genuine videos and 49 fake videos having around 32752 frames in total. Compared to previous methods that use a large amount of real and DeepFake images to train CNN classifier, this method doesn't need any negative training examples as it targets the objects in affine face warping as the distinctive feature for distinguishing real and fake images. This method was evaluated on the UADFV dataset on “VGG16” [26], ResNet50, ResNet101, and ResNet152 [27] models using the Area Under Curve (AUC) metric in two settings: image-based evaluation and video-based evaluation. For image-based evaluation, they process and send frames of all videos into the four networks respectively. The VGG16, ResNet50, ResNet101 and ResNet152 models achieve AUC performance 83.3%, 97.4%, 95.4%, 93.8%, respectively. ResNet networks have about 10% better performance compared to VGG16, due to the residual connections, which make the learning process more effective. Yet, ResNet50 has the best performance among the other ResNet networks. The video level performance of each type of CNN model. VGG16, ResNet50, ResNet101 and ResNet152 can achieve AUC performance 84.5%, 98.7%, 99.1%, 97.8%

respectively. In this video-based evaluation metric, the ResNet network still performs ∼15% better than VGG16. Yet, each ResNet model has a similar performance, as in the case of image-level classification.

- **X. Yang, et al. [11], 2019,** proposed a new way of revealing DeepFake videos created by Artificial Intelligence methods. This method is based on the observations that DeepFakes is created by splicing the synthesized face region into the original image, thereby introducing errors that can be revealed when the face images estimate 3D head poses. An SVM classifier is evaluated using a collection of real face images and DeepFakes using features based on this cue. The results, assessed using individual frames as an inspection unit with the output metric Area Under ROC (AUROC). The tests show the SVM classifier reaches an AUROC of 89.0% on the UADFV dataset. This indicates that the estimated difference between the head and the whole face from the central region is a good feature for identifying images generated by DeepFake. Besides, an estimation of the performance was carried out using individual videos as an analysis unit for the UADFV dataset. This is done by averaging the estimation of frame classification over the individual videos. They also conduct an ablation study comparing the performance of various types of features used in the SVM classifier.
- **E. Sabir et al. [28], 2019,** proposed a method that exploits the Spatio-temporal highlights to distinguish DeepFakes videos. Recurrent convolutional layers models are a class of profound learning models that have proved effective in exploiting temporal information from domain-wide image streams. Thus the best strategy for combining variations in these models with domain-specific face preprocessing techniques is refined through extensive experimentation to obtain state-of-the-art performance on benchmarks of publicly available video-based facial manipulation. Specifically, the attempt is to identify tampered faces in video sources achieving AUC results of 96.9% and 96.3% for the DeepFake and FaceSwap methods, respectively. Only the low-quality

videos were considered in the analysis. The test is also carried on the FaceForensics++ dataset [29], improving the previous state-of-the-art by up to an accuracy of 4.55% to reach 94.3% ACC.

- **H. Nguyen, et al. [30], 2019,** proposed detecting manipulated videos by the use of Capsule Networks. This type of network was introduced at the beginning to address the CNNs limitations when used in inverse graphics tasks. The evaluation of this method was on four datasets having a wide range of fake videos and images, which include the Ldiap Research Institute replay attack dataset [31]. The accuracy of face swapping detection at frame level on the DeepFake dataset was 95.93% and the accuracy of face swapping detection at video level on the DeepFake dataset was 99.23%.
- **O. de Lima et al. [32], 2020,** Showed that intra-frame inconsistency and temporal inconsistency among frames are found in DeepFake videos. A temporal-aware pipeline method was proposed which utilize CNN and Long Short Term Memory (LSTM) to spot DeepFake videos. Frame features are extracted by the CNN, after that it is passed into the LSTM to generate a descriptor of the temporal sequence. Afterward, for classifying manipulated videos from genuine videos, a fully-connected network is used based on the sequence descriptor. The system using the Celeb-DF dataset can accurately predict if the fragment being analyzed comes from a DeepFake video or not, this method outperformed state-of-the-art frame-based detection methods. This method tested some of the most popular networks that take advantage of temporal features. All the networks were trained on the Celeb-DF dataset starting from the pre-trained published weights. No layers were frozen for training. Each method was trained for over 25 epochs and the best ROCAUC scores were 74.87%, 99.43%, 97.59%, 99.30%, and 99.73 on Residential Communications Network, R2Plus1D, I3D, Mobile CubeSat Command and Control (MC3), and R3D respectively.

1.4 Statement of the problem

DeepFake technology can learn and utilize from the massive amounts of photos and videos found on the Internet to generate not only forged videos but in a very hyper-realistic video of individuals. Scampers may use these renderings to target different types of public figures and political leaders, such as the presidents of the leading countries, even executives of major celebrities. The consequence of this phenomenon would lead to distrust in what to hear or what to see. DeepFakes are a seemingly realistic video of an individual's generated using artificial intelligence that shows actions that never occurred in reality. These types of videos are becoming almost indistinguishable from the real videos.

Now, as we are living in the digital age, the ability of denial and fraud campaigns have advanced more than ever before, by coordinating online campaigns to spread artificial false, misleading, or malignant content. The habit of creating an alternative reality is a result of the incompetence of people to believe what they hear or see, people are more likely to pick out the reality that most mightily aligns with their thought. The substitutional reality may divide society by engaging in people's prejudice and by eliminating the common understanding of the truth.

1.5 The aim of the thises

This thesis aims to build a strong DeepFake detection system to distinguish between real and fake videos that have been generated by DeepFakes applications by using Discrete Wavelet Transform (DWT). Besides the use of DWT, a pre-trained ResNet-50 has been used to boost the capture of the artifacts to give more accurate results. Serious steps must be taken, and start developing tools for the detection of DeepFakes videos to limit the creation of this phenomenon

1.6 Thesis organization

The thesis is segmented into five chapters; a brief description of their contents is given below:

Chapter One: This chapter introduces an overview of the work and related works.

Chapter Two: This chapter introduced methods and descriptions for the theoretical background and techniques that are used in this thesis.

Chapter Three: This chapter describes the proposed systems with their design and implementation and the execution of the stages of the proposed system.

Chapter Four: This chapter presents the tests and the results of the proposed system.

Chapter Five: This chapter offers conclusions and systems for future work.