Banknote Authentication and Medical Image Diagnosis Using Feature Descriptors and Deep Learning Methods

Tamarafinide Victory Dittimi

A Thesis
In the Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Computer Science) at
Concordia University
Montreal, Quebec, Canada

May 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Tamarafinide Victory Dittimi

Entitled: Banknote Authentication and Medical Image Diagnosis Using Feature Descriptors and Deep Learning Methods

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair

Dr. Lius Amador

_____External Examiner

Dr. Mounim El-Yacoubi

_____Examiner

Dr. Adam Krzyzak

_____Examiner

Dr. Olga Ormandjieva

_____Examiner

Dr. Yong Zeng

_____Supervisor

Dr. Ching Suen

Approved by _____

Dr. Volker Haarslev, Graduate Program Director

June 28, 2019 _____

Dr. Amir Asif, Dean

Faculty of Engineering and Computer Science

ABSTRACT

Banknote Authentication and Medical Image Diagnosis Using Feature Descriptors and Deep Learning Methods

Tamarafinide Victory Dittimi, Ph.D.

Concordia University, 2019

Banknote recognition and medical image analysis have been the foci of image processing and pattern recognition research. As counterfeiters have taken advantage of the innovation in print media technologies for reproducing fake monies, hence the need to design systems which can reassure and protect citizens of the authenticity of banknotes in circulation. Similarly, many physicians must interpret medical images. But image analysis by humans is susceptible to error due to wide variations across interpreters, lethargy, and human subjectivity. Computer-aided diagnosis is vital to improvements in medical analysis, as they facilitate the identification of findings that need treatment and assist the expert's workflow. Thus, this thesis is organized around three such problems related to Banknote Authentication and Medical Image Diagnosis.

In our first research problem, we proposed a new banknote recognition approach that classifies the principal components of extracted HOG features. We further experimented on computing HOG descriptors from cells created from image patch vertices of SURF points and designed a feature reduction approach based on a high correlation and low variance filter.

In our second research problem, we developed a mobile app for banknote identification and counterfeit detection using the Unity 3D software and evaluated its performance based on a Cascaded Ensemble approach. The algorithm was then extended to a client-server architecture using SIFT and SURF features reduced by Bag of Words and high correlation-based HOG vectors. In our third research problem, experiments were conducted on a pre-trained mobile app for medical image diagnosis using three convolutional layers with an Ensemble Classifier comprising PCA and bagging of five base learners. Also, we implemented a Bidirectional Generative Adversarial Network to mitigate the effect of the Binary Cross Entropy loss based on a Deep Convolutional Generative Adversarial Network as the generator and encoder with Capsule Network as the discriminator while experimenting on images with random composition and translation inferences. Lastly, we proposed a variant of the Single Image Super-resolution for medical analysis by redesigning the Super Resolution Generative Adversarial Network to increase the Peak Signal to Noise Ratio during image reconstruction by incorporating a loss function based on the mean square error of pixel space and Super Resolution Convolutional Neural Network layers.

ACKNOWLEDGEMENTS

It is indeed the greatest joy and pride in my life to write this acknowledgment for my doctoral dissertation. The overwhelming sense of accomplishment brings warmth to my heart, which I would like to share with those who made it possible for me to embark on this journey and stood by me at every step of the way. I would like to express my gratitude to all those who have supported, influenced and helped me in the process which ultimately resulted in this thesis.

First and foremost, I humbly express my sincere and profound gratitude to my supervisor and mentor, Dr. Ching Yee Suen, who, for many years, was the guiding light in my graduate studies. This dissertation would have never been possible without his invaluable guidance, insightful feedback, remarkable patience, and meticulous editing. I am forever indebted to him for seeing more in me than I saw in myself. Thank you for teaching me that no dream is ever too big to pursue. I would also like to thank Mr. Nicolas Nobile and Dr. Marleah Blom for their support and encouragement. Also, I am also grateful to my committee members, Dr. Adam Krzyzak, Dr. Olga Ormandjieva, and Dr. Yong Zeng for their valuable comments and suggestions.

Furthermore, I would also like to thank all my colleagues at Concordia University especially the members of the Centre for Pattern Recognition and Machine Intelligence for their fruitful comments and discussions during my research work and this dissertation. I am very grateful to all other friends who helped me out with all the challenges of living in Canada and always supported me.

It is also my absolute honor and duty to thank my parents, who generously and wholeheartedly gave me their unconditional love and endless support throughout these years. This dissertation pales into insignificance compared to the sacrifices they made for me to make it this far. I am grateful for your trust and confidence in me and for giving me the freedom to pursue my dreams. I, forever, will treasure your love in my heart. This work is also dedicated to my wife, Faith, and my lovely kids Dior and Johnson for their unwavering love and encouragements during the pursuit of my studies. Thank you for always believing in me, for accompanying me on those long nights of writing, and for reminding me to endure during the tough times. I am genuinely thankful for having all of you in my life. Lastly, I would like to dedicate my thesis to my father Dr. Johnson Dittimi who although isn't here to see me complete the program, would be overjoyed by my completion of the Ph.D. Program.

CONTRIBUTION OF AUTHORS

This section lists all the papers that I previously published or submitted in international conferences and journals (including co-author papers) since I started my Ph.D. program at Concordia University.

International Conferences (Published)

- **T.V. Dittimi**, A.K. Hmood, and C.Y. Suen (2017, March), Multi-class SVM based Gradient feature for banknote recognition, In Proc. of the International Conference in Industrial Technology (ICIT), Toronto, pp. 1030 – 1035.
- A.K. Hmood, **T.V. Dittimi**, and C.Y. Suen (2017), Scale and Rotation Invariant Character Segmentation from Coins, In Proc. of the 14th International Conference on Image Analysis and Recognition (ICIAR), Montreal, pp. 153 – 162, LNCS, Springer.
- **T.V. Dittimi**, A.K. Hmood, and C.Y. Suen (2018), Mobile Based assistive Technologies for Authentication of Banknote, In Proc. of the International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI 2018), Montreal, pp. 366 – 371.
- A.K. Hmood, **T.V. Dittimi**, and C.Y. Suen (2018), Counterfeit Coin Detection using Stamp Features and Convolutional Neural Network, In Proc. of the International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI 2018), Montreal, pp. 273 – 278.
- **T.V. Dittimi**, and A.K. Hmood (2018), Mobile App for Detection of Counterfeit Banknote. In Proc. of the Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, 31, Toronto, pp. 156-168.

Journal Papers (Published)

- **T.V. Dittimi** and C.Y. Suen (2018), Modified HOG Descriptor-Based Banknote Recognition System. Adv. Sci. Technol. Eng. Syst. J. Vol. 3, Issue 5, Pages 354-364.
- **T.V. Dittimi** and C.Y. Suen (2018), High correlation-based banknote gradient assessment of ensemble classifier. International Journal of Wavelets, Multiresolution and Information Processings, Pages 1950006-1950012.
- **T.V. Dittimi** and C.Y. Suen (2019), Mobile Phone based ensemble classification of Deep Learned Feature for Medical Image Analysis. IETE Technical Review, Pages 1-12.

Table of Contents

List of Figures

List of Tables

List of Acronyms

1. ADDA - Adversarial Discriminative Domain Adaptation
2. ATM - Automated Teller Machine
3. BIGAN - Bidirectional Generative Adversarial Network
4. BOW - Bag of Words
5. BPNN - Back Propagation Neural Network
6. CAD - Computer-aided diagnosis
7. CNN - Convolutional Neural Network
8. DCGAN - Deep Convolutional Generative Adversarial Network
9. DT - Decision Tree
10. ECOC - Error Correcting Output Code
11. GAN - Generative Adversarial Network
12. GGD - Generalized Gaussian Density
13. GLCM - Gray Level Co-occurrence Matrix
14. GLOH - Gradient Location and Orientation Histogram
15. HCF - High Correlation Filter
16. HOG - Histogram of Gradient
17. HWT - Haar Wavelet Transform
18. KNN - K-Nearest Neighbor
19. LBP - Local Binary Pattern
20. LDA - Linear Discriminate Analysis
21. LVF - Low variance filter
22. LVQ - Linear Vector Quantization
23. MCSVM - Multi-Class Support Vector Machine
24. MLPNN - Multi-Layer Perceptron Neural Network
25. NB - Naïve Bayes
26. OVOSVM - One-Versus-One Support Vector Machine
27. PCA - Principal Component Analysis
28. QWT - Quaternion Wavelet Transform
29. RF - Random Forest
30. RNN - Recurrent Neural Network
31. ROC - Receiver Operating Characteristic
32. RVM - Relevance Vector Machine
33. SC - Shape Context
34. SIFT - Scale Invariant Feature Transform
35. SRCNN - Super Resolution Convolutional Neural Network
36. SRGAN - Super Resolution Generative Adversarial Network
37. SSL - Semi-Supervised Learning
38. SURF - Speeded-up Robust Features
39. SVM - Support Vector Machine
40. WHO - World Health Organization
41. WMA - Weighted Majority Average

Chapter 1

1 Introduction

1.1 Background

Technological advancement has made life easier but has its downside too, as criminals have also taken advantage of this progression. Initially, only printing houses had access to the technology needed to produce fake banknote but today with the aid of a personal computer and a good laser printer people can duplicate paper money easily in their residence. In this age of scientific advancement, automated machinery for handling banknote plays a significant role in our daily life as financial institutions perform thousands of monetary transactions every day. Furthermore, the improvement in modern-day banking operations and automated systems for identification of currencies have become pertinent in many applications such as electronic banking, currency monitoring system, money exchange machine, Automated Teller Machine (ATM) and vending and parking meters [88].

Counterfeiting is the manufacturing of replicated currency without the lawful authorization of the state or government, usually done for fraudulent purposes. It has been in existence since the coinage of money started in 600 B.C. Greece. At that time, coins were being utilized as legal tenders. It was not until the 1200s that paper currency was invented in China using the wood from mulberry trees. At that time counterfeiting was prohibited by the Chinese guards securing mulberry forests and preventing access by enforcing the death penalty on any individual caught trying to produce banknote illegally. However, even in modern times, this problem is still evident thus the need to employ the extrinsic or intrinsic properties of the banknote to prevent and possibly avoid note forgery [98].

Smartphone usage has become part of our daily life and their development is still snowballing. Most phones possess advanced functionalities that go beyond making calls and text messaging.

They also serve as portable computing devices used in running e-mails, agendas, and storing data. Furthermore, mobile devices have at least one high-resolution camera that can be used to capture images and also hold sufficient computational capability to process the image and execute algorithms for the recognition of banknote [110]. People identify currencies using several methods like inspecting the pictures, words, and numerals as well as distinctive signs, note consistency, and size. However, these techniques are prone to human error or omissions caused by visual impediments. According to the World Health Organization (WHO), there were about 285 million visually impaired people worldwide, of which 39 million were visually impaired, and 246 million had low vision. Furthermore, one of the most significant problems faced is the ability to detect the value of a banknote [47], since most people differentiate banknote denominations by folding them in different distinct modes. Although this works with banknote already in their possession, it would be difficult for them to identify note received during daily transactions.

Physicians perform most interpretations of medical images; but, image analysis by humans is susceptible to error due to wide variations across interpreters, lethargy, and human subjectivity. Computer-aided diagnosis (CAD) is vital to improvements in medical analysis, as they facilitate the identification of findings that need treatment and assist the expert's workflow [11]. Several researchers have proposed incorporating CAD into the medical diagnostic process. It is a computerized procedure that offers a quantitative opinion to improve the clinical diagnostic process by providing an alternative view of medical image interpretation and diagnosis. Surely, CAD has proven to be useful to assist in the reduction of unnecessary false-positive diagnosis, biopsies, and thoracotomy [103].

1.2 Problem Statement

Banknote Identification is a relatively easy job for human eyes as the brain is capable of absorbing different information and recognizing them with less effort. In image recognition, identification of banknote is a challenging task as notes can suffer from many defects once in circulation. Also, during the process of acquiring currencies via digital camera or scanners, some images become skewed due to alignment and acquisition setup [79]. Presently, no country is exempted from the problem faced because of the counterfeiting of its banknote. It is a significant issue that can cripple a country's economy, since note forgery increases the circulation of paper

currency leading to higher inflation and devaluation of the money, reduce people savings, loss in national economy and public confidence [5]. Furthermore, counterfeit Nigerian currencies are very much in circulation in Nigeria and although the regulatory body does not have a record of the incidences of fake banknote that are in circulation since the deposit-taking institutions which intercept such note do not keep the records and report same. It is still a fact that counterfeit currencies are very much in circulation in Nigeria today, and should be a matter of great worry for the government because of the grave consequences they tend to produce on the economy [26].

Over the years, differentiating between fake and real currency banknote has become quite tricky due to the fast-scientific progress in color printing, cloning, and imaging and because fake note is now reproduced via top-notch technology that uses security paper. In the same view, counterfeit banknote detection has become an essential task in billing machines. It is implemented by using image analysis methods (transmittance analysis, reflectance analysis) with different light spectra (visible, infrared and ultraviolet). The drawback of this system is its high cost and its usage with only ATM, where an elevated level of reliability is required [3].

Mobile phone usage has become very common, and its market is still growing. Most mobile phones now have advanced functionalities that go beyond making phone calls and sending text messages [7]; they are a portable computing device that is used in sending-mails; create agendas and store data. Also, Smartphones have at least one high-resolution camera that can be used to capture images and hold enough computational capability to process the images and execute algorithms for our banknote recognition system [110].

Visually impaired people are confronted with many challenges during their daily lives, especially conducting financial transactions. Although several traditional techniques exist that assist in identifying the different denominations of the banknote, they also have their problems and limitations to its effectiveness. However, over time note get old and the lines could get blurry making it difficult for visually impaired to accurately identify the different note denominations. Another option employed was differentiating banknote denomination by folding them into different distinct modes, while this works with bills already in possession it is still difficult to identify paper money received during daily transactions. Also, many currencies possess tactile marks used by visually disabled people in further differentiating the denomination, but note gets worn out during circulation which could affect the Braille marks. Although several security features have been

incorporated into banknote; counterfeiters have capitalized on the innovation in printing media technologies in reproducing fake paper currencies.

The field of medical pathology and radiology is tasked with providing absolute disease diagnoses to support and direct patient treatment and managerial decisions. Furthermore, the implementation of a CAD system to standardize the accuracy of diagnoses to ensure reproducibility that is essential for advancing precision in medical diagnosis has always been imperative. For over 100 years, the main approach employed by pathologists for medical diagnosis was microscopes before advancements led to the invention of x-rays [35]. However, human interference based on visual analysis of radiology and microscopic images has been plagued by the lack of standardization, diagnostic blunders, and the substantial cognitive load required to evaluate the images manually. Consequently, in the last two decades, several researchers have seen the need to design computational techniques that can assist in the analysis of medical pathology and radiology images [38].

Thus, deep learning systems have been introduced into the medical image analysis domain with promising results on various applications, like a computerized diagnosis of breast cancer, mild cognitive impairment, blood cell classification, bone age estimation, Alzheimer's disease, radiology, organ segmentation and detection, and ultrasound standard plane selection. Thus, deep learning is rapidly proving to be the state-of-the-art system, leading to improved accuracy thus opening new frontiers in medical analysis with rates of progress not experienced before [106].

CAD systems have focused primarily on machine learning methods using image analysis methods like normalization, segmentation, and feature extraction. After which classification models are constructed using standard machine learning methods including Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), K-Nearest Neighbor, and Decision Tree (DT) or deep learning based systems like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Generative Adversarial Network (GAN) with no discrete human-directed steps for segmentation, detection, and feature extraction. Instead, the deep learning algorithms take as input only the images and its corresponding labels and train a very high-dimensional and complex set of model parameters with supervision coming only from the image labels [61]. Deep neural networks have achieved commendable performance in specific supervised learning problems by leveraging large collections of labeled data. However, these successes come at a distinct cost; involving the creation of large dataset which requires a great deal of human effort. An attractive approach to

mitigating this issue led to the introduction of Semi-Supervised Learning (SSL). It is halfway between supervised and unsupervised learning. A small bit of labeled data is provided to the systems and serves as the starting point for the computer systems and the systems accept and learn from large volumes of unlabeled data.

1.3 Research Aim

The main aims of this research are as follows:

1. Design a framework to ease the process of identification of banknote denominations, counterfeit paper currency detection and Mobile phone-based identification of original or forged banknotes to aid the visually impaired.
2. Propose a medical image diagnosis system using deep learning methods to handle recovering any high-frequency information, segmentation, normalization, random composition, translation inferences and conduct continuous recursive self-improvement in a Semi-supervised setting.

1.4 Research Objectives

1. This research will propose a new banknote recognition approach that classifies the principal components of extracted HOG features and would further experiment on computing HOG descriptors from cells created from image patch vertices of SURF points before designing a feature reduction approach based on a high correlation and low variance filter.
2. We would develop a mobile app for banknote identification and counterfeit detection using the Unity 3D software and evaluate its performance based on a Cascaded Ensemble approach.
3. Furthermore, the banknote recognition algorithm will then be extended to a client-server architecture using SIFT and SURF features reduced by Bag of Words and high correlation-based HOG vectors.
4. An experiment would be conducted on a pre-trained mobile app for medical image diagnosis using three convolutional layers with an Ensemble Classifier comprising PCA and bagging of five base learners.
5. The research would also implement a Bidirectional Generative Adversarial Network to mitigate the effect of the Binary Cross Entropy loss based on a Deep Convolutional Generative Adversarial Network as the generator and encoder with Capsule Network as the discriminator while experimenting on images with random composition and translation inferences.

6. Lastly, we will propose a variant of the Single Image Super-resolution for medical analysis by redesigning the Super-Resolution Generative Adversarial Network to increase the Peak Signal to Noise Ratio during image reconstruction by incorporating a loss function based on the mean square error of pixel space and Super-Resolution Convolutional Neural Network layers.

1.5 Research Contribution

The main contributions of this thesis are as follows.

1. We extended the global-based feature method by exploring the identification of images using the computation of the Principal Component Analysis (PCA) of the Histogram of Gradient (HOG) function and introduced a variant of the Multi-Class Support Vector Machine (MCSVM) that uses an error correcting output code algorithm to improve the classification results of the banknote.
2. We presented a cellphone-based banknote recognition application that extracted the face value of the banknote and evaluated its performance using a combination of several KNN distant measures based on a Cascaded Ensemble approach determining their authenticity.
3. We recommended a client and server-based mobile application for banknote recognition system that extracts a pixel feature based on the Histogram of Gradient; then reduces the feature set using PCA, Bag of Words (BOW), Low Variance, and High Correlation feature reduction technique. Lastly, the classification was done using the Weighted Majority Average for combining Multi-Layer Perceptron Neural Network and Multi-Class Support Vector Machine.
4. We designed a keypoint based descriptor system that employs a client-server architecture where the mobile app built from Unity 3D captures the paper money. The image is preprocessed before transmitting to a MATLAB system that uses a three-level algorithm to extract and combine SIFT and SURF features. The feature set is reduced by Bag of Words to classify the bills using a stacked MLP and ECOC method.
5. We also exploited deep learning techniques for medical image analysis on mobile phones; the system comprised of 3 convolutional layers using several filters with varying dimensions containing two max-pooling and batch normalization layers. The Relu layer was implemented in all the Convolutional Networks, and the learned feature output is extracted using the fully

connected layers based on nodes constructed at each layer. While the Ensemble Classifier consists of a PCA based feature reduction, and five base learners using bagging to classify medical image dataset.

6. We also systematically study GANs for imaging Analysis using SSL, implemented a Bidirectional Generative Adversarial Network (BIGAN) architecture to mitigate the effect of the BCE loss because GAN has a perceptual loss function and introduced Deep Convolutional Generative Adversarial Network (DCGAN) autoencoder by rearranging its structure for a generator and encoder architecture for BIGAN.
7. We designed a GAN discriminator architecture by extending the Capsule network that replaces the max-pooling with routing dynamic routing mechanism and the scalar-output feature detectors with a vector-output Capsule. But it still utilizes translated replication of the trained feature detectors achieving reproducibility of learned knowledge across space and uses a convolutional layer in all the last layers and experimented using images with random composition and translation inferences in a Semi-Supervised setting by proposing a paradigm of modular learning which conducts continuous recursive self-improvement with regards to the utility function (reward system) viewing it this as second and higher order optimization method.
8. We proposed a variant of the Single Image Super-resolution for medical analysis by redesigning the SRGAN. By using VGG19 network for feature extraction, setting discriminator network's working space as feature space, and adding the loss function based on the mean square error of pixel space, gaining more details by incorporating SRCNN layers to increase the PSNR in the reconstruction at the same time.

1.6 Thesis Outline

The thesis is organized as follows:

- Chapter 2: We discussed the preliminary knowledge and reviews of related works including banknote identification and authentication methods like texture-based recognition, country-specific banknote recognition, feature descriptor-based banknote recognition, counterfeit detection, and mobile-based recognition system. On the other hand, we also evaluated computer-aided medical image diagnosis using the convolutional

neural network and generative adversarial network for photo-realistic image generation and single image super-resolution.

- Chapter 3: We proposed a new banknote recognition approach that classifies the principal components of the extracted histogram of gradient feature vectors using an efficient error correcting output code technique based on a multi-class support vector machine. We further experimented on a banknote recognition and counterfeit detection system that utilizes a pre-processing approach before creating a new feature set by extracting Speeded-Up Robust Feature, creating image patches from its vertices and then computing the Histogram of Gradient descriptors from cells within the patch boundaries. Before a feature reduction approach is also designed based on a high correlation and Low variance filter and classification performed using 10-fold cross-validation based One-Versus-One Support Vector Machine (OVOSVM).
- Chapter 4: We developed a mobile app for the banknote identification and counterfeit detection using Unity 3D. The system extracted the banknote face value and evaluated its performance using a combination of several KNN based on a Cascaded Ensemble approach. The algorithm was then extended to a client-server architecture that combines SIFT and SURF features before reduction by Bag of Words to classify the bills by using a stacked MLP and ECOC on a MATLAB backend. Additional investigation was also conducted to introduce new approaches for recognition and authentication of banknotes; the system extracted the shape context, Scale Invariant Feature Transform, gradient location and orientation histogram, and Histogram of Gradient. It then reduces the feature set using PCA, BOW and proposed two-dimension reduction approach based on low variance and high correlation filter. The classification was done using a 2-fold Weighted Majority Average Ensemble technique with MPLNN and MCSVM as base classifiers.
- Chapter 5: We proposed a pre-trained mobile application for medical image diagnosis; it examined the benefit of deep learning approaches for White blood cell and chest radiography analysis. The feature extraction network comprised three convolutional layers using several filters with varying dimensions containing two max-pooling and batch normalization layers. The Relu layer was implemented in all the Convolutional Networks, and the learned feature output is extracted using the fully connected layers based on nodes constructed at each layer. While the Ensemble Classifier consists of a Principal Component

Analysis based feature reduction, and five base learners using bagging to classify medical image dataset.

- Chapter 6: We implemented a Bidirectional Generative Adversarial Network that achieved significant improvements in generating photo-realistic medical images using a semi-supervised method that mitigates the common problems of global consistency, random dimension and translation, mode collapse, and stabilizing training of GANs. It extended the architecture to mitigate the effect of the BCE loss using a DCGAN as the generator and encoder with Capsule Network as the discriminator while experimenting using images with random composition and translation inferences.
- Chapter 7: We propose a variant of the Single Image Super-resolution for medical analysis by redesigning the SRGAN. By using VGG19 network for feature extraction, setting discriminator network's working space as feature space, and adding the loss function based on the mean square error of pixel space, gaining more details by incorporating SRCNN layers to increase the PSNR in the reconstruction at the same time.
- Chapter 8: We presented a summary of the thesis' contributions and discussions for possible future developments of the proposed Banknote authentication and Medical Image analysis.

Chapter 2

2 Literature Review

2.1 Banknote Identification and Authentication

Banknotes are utilized in every business venture in as much as electronic currencies is becoming identified as a substitute legal tender, the note is still dominant and central in the day to day dealings. Currency recognition and detection is a critical problem in computer vision. Although electronic and online currencies exist, banknote is still the preferred option in daily transactions such as vending machine, the banking system, banks, shopping mall, railway ticket counter, ATM, vending machines, and currency exchange service. Furthermore, there are cases of money bundles withdrawn from ATMs contained a few counterfeit bills although banks are unwilling to accept these omissions. Even though this problem is unavoidable, they are not able to provide all branches with the needed counterfeit detection system [118].

2.1.1 Texture-Based Banknote Recognition

One of the most straightforward methods that have been applied in paper money recognition system is the use of visible features found on banknotes like the size, texture, and color of banknote. All banknotes have pictures and patterns impressed on them, such as portraits of historical persons or images of landmark buildings which are characterized as textures extracted as features using texture-modeling techniques. The drawback of this approach is that banknotes are damaged and torn due to regular use, leading to alteration of the note thereby hurting the recognition rate [48]. On the other hand, [58] extracted the geometric features (length and width) and employed it in classifying bills with varying sizes such as Italy and Chinese banknote. Although, this approach was simple and fast; it had significant limitations including its effectiveness for slant, stain, and tear. Furthermore, it was unsuitable for some banknotes such as the US Dollars, which had several denominations of the same size.

Furthermore, many countries employ colors for differentiating banknote. Despite wear and tear, they still retain their color [37]. [2] designed a system that identified the denomination of the Indian currency note based on its color. Likewise, another author also developed a technique that converted the image from RGB color space to HSV color space which then extracted the color feature. Their justification for employing the HSV color space was because of its nearness to the human conceptual understanding of color [87]. Additionally, a researcher recommended an approach based on the characteristic point extraction of a unique part of the Euro paper currency that possesses the same color has also been conducted. In the same view, [67] presented research on coordinate data extraction approach based on the characteristic point extraction of the unique part of the Euro paper currency that possesses the same color. Similarly, [21] designed a method for the classification of Korean Won bill, the system acquired test images using a visible light spectrum. Images were processed using a Sobel filter, and its wavelet domain features were extracted. Furthermore, the canonical analysis projection was employed to reduce feature space dimensionality and the K-nearest neighbor method used for final banknote recognition.

2.1.2 Country Specific Banknote Recognition

Several methods have been designed for identifying and authenticating banknote of specific countries. A Canadian banknote reader has been developed, but the system only identified Canadian note and its denomination by outputting a pre-recorded voice output informing the user of the note value [24]. Furthermore, a system for recognition of Sri Lankan currency was also designed by [44], the method adopted a unique linear transformation function for noise reduction, edge detection for fair representation of boundaries for new and old damaged notes, and a three-layer back propagation Neural Network for Classification.

Additionally, [121] proposed an Indian currency recognition system based on the intrinsic features of banknote image; the research was conducted on the front side of 20 note images in denominations of Rs 10, 100, & 500. Similarly, [82] proposed a general object reader called EyeRing that was trained to recognize the value of USA paper money. It was connected to a mobile phone through the Bluetooth facility and employed in identifying the value of the US dollar. Though it had its limitations regarding the hardware and technical know-how needed to utilize the system. Also, a compact device for recognizing US Dollar denomination has been developed; the system is called the iBill Talking Banknote Identifier. It utilizes a combination of image processing and

feature extraction method to distinguish the note and then uses a pre-recorded voice output, sequence of tones and vibrations to inform the user of the banknote value. The limitation of this approach is that it works on only US dollar bills and cannot be employed to detect counterfeit US note [84].

In the same view, [118] recommended an image-based Bangladesh paper money recognition technique, the system extracted the Histogram of Gradient (HOG) feature from the watermark, Latent Image, and micro-printing of the note and classified it using an SVM method. Finally, the approach had a 100% recognition rate. Despite, it was only tested on the 500 BDT and 1000 BDT. Likewise, [134] developed an Ethiopian Birr note recognition and detection approach, the paper money was captured using a scanner consisting of Ethiopian and non-Ethiopian note including namely US dollar, Rand, Dirham, and Shilling. The Birr note comprised of real and fake 100 Birr note, 50 Birr note, 10 Birr note, 5 Birr note, and 1 Birr note. The technique classified the bills into one of the Ethiopian denominations using the dominant color, distribution of dominant color, hue value and SURF matching. Furthermore, for authentication of the note, the thin and wide strip on the banknote was utilized. Finally, the system had an average denomination recognition rate of 90.42% with an average counterfeit rejection rate of 100%. On the other hand, the technique adopted for authenticating the banknote is flawed as it is only functional with a high-quality note that retains the number of objects in thin and wide strips.

Also, [114] suggested a Thai currency recognition system which extracts the slab values and feeds it to a Neural Network to achieve its learning and recognition process. Comparably, a method that improves the recognition rates of banknote using principal component analysis (PCA) was also devised by [4]. The system was tested on six different bill-types of US Dollars; the data was acquired using a line sensor, the PCA algorithm was used to extract features and normalize the data size while a Linear Vector Quantization (LVQ) network was employed for classification. The result revealed that when the number of PCA components and the number of LVQ codebook vectors are taken correctly; the reliability could be increased.

2.1.3 Feature Descriptor Banknote Recognition

Several scholars have proposed numerous global-based recognition systems; [96] employed a technique for the classification of Slovak banknote based on SIFT detector and descriptors.

However, the system used a server-client model and had a long recognition time. At the same time, a Euro recognition system, also based on SIFT, was proposed. This method utilized the bag of words scheme and classified the banknote using SVM, ANN, and HMM [94]. Furthermore, [20] addressed a side invariant technique for the recognition of Indian paper currency that applied a template matching method. The system compared the edges of the test image and matched them to the edges of the reference image stored as templates in a database. However, the system has its flaws that include a low recognition rate, lighting illumination, image quality, and positioning of the banknote during the image acquisition. Furthermore, [44] also described a method for recognition of Sri Lankan banknote. This approach adopted a unique linear transformation function for noise reductions and edge detections for a good depiction of the edges of new and old damaged note. The currencies were classified using a three-layer Back Propagation Neural Network (BPNN). However, the method is side variant and unsuccessful on partial banknote.

Another paper also presented a Chinese Renminbi banknote recognition method using the Block Local-Binary Pattern. The method was robust to noise and illumination change [45]. Additionally, [36] designed a multi-currency banknote system and tested the approach using Chinese Renminbi and Euro note. The system used the Quaternion Wavelet Transform (QWT), Generalized Gaussian Density (GGD) and a Back Propagation Neural Network (BPNN) method. [3] also discussed a Persian currency-based approach. The system employed the Haar Wavelet Transform (HWT) and Multi-Layer Perceptron (MLP) neural network. Although wavelet transform is a powerful tool, this method is shift sensitive as a change in the input signal can lead to an unpredictable change in the transform coefficient. The system also had other flaws such as with banknote feature direction, as the method only caters for a spatial-domain feature in a horizontal, vertical and diagonal orientation.

A researcher [71] further asserted the need for a note classification method by proposing a system based on grid features. The system divided each banknote into m x n partitions and used the average gray value of each partition as its feature vector. However, the technique has its setback as it had issues in determining the best size and number of features for efficient classification. Similarly, a feature-based recognition algorithm based on SURF was studied. The system matched reference regions with query images. Its proposed process achieved a high recognition rate on the dataset and could recognize worn or wrinkled bills [47]. It was only tested on US dollars and employed a small dataset. Besides, the system had high computational time and extracted SURF

features from specific fixed locations in the banknote. Lastly, [50] presented an approach for recognition of Nigerian paper currencies using SVM and a genetic algorithm. The system used a genetic algorithm for parameter optimization, and the note was classified using an SVM-based classifier on a predefined decision function. However, the technique was not further evaluated using other global currencies and did not present an efficient set of features for classification.

2.1.4 Mobile Based Banknote Recognition

Several mobile phone-based banknote recognition and detection systems have been designed and reported to aid visually impaired individuals. [108] Some of these methods include LookTel Money Reader and IDEAL Currency Identifier. Both of which are intended to recognize any banknote. Country-specific applications have also been developed like the Money Talker for Australian dollar recognition along with Note Teller 2 and K-NFB mobile reader, both designed for US dollar recognition. Similarly, a camera phone-based banknote recognition system was also proposed to identify the denomination of a U.S note [69]. It utilized a background subtraction and perspective correction algorithm and trained the currency reader using an efficient Ada-boost framework. The system works on some of the leading smartphone platforms like Symbian and Windows Mobile. However, the system is yet to be evaluated with state-of-the-art methods and the recognition time needs to be improved even though the system works in real time. Furthermore, another research also employed a real-time portable object recognition system. The system is based on the bio-inspired image analysis of visual cues to increase the autonomy of blind people. The method was built to recognize the 5€, 10€, 20€ and 50€ bills. The device could detect banknote regardless of its orientation. However, the note was captured using a webcam and processed on an Ultra Mobile PC using Spikenet and had yet to be adapted to run on mobile phones. Likewise, developed a technique for Myanmar currency recognition, which was based on the KNN classification of extracted Gray Level Co-occurrence Matrix (GLCM). The dataset used consisted of 500 images grouped into five classes of banknote. The limitation of the system is its poor accuracy in determining the denomination of real-world Myanmar note. Correspondingly, [26] described a technique for recognizing the US Dollar using an image recognition approach called Eigenfaces based on the Principal Component Analysis (PCA). The system was tested on two different Nokia Smartphones, and it had a precision rate of 99.8% and a processing speed of at least seven frames per second. The system faced issues with the image acquisition illumination

and differentiating the background from the banknote. Likewise, another researcher suggested an approach for the blind that was built to run on smartphones to identify note like the Euro, Zloty and the US Dollar denominations. The method used computation based on the height and width, which is not ideal because they change due to wrinkling, wear, and tear of the banknote.

Similarly, another work introduced a paper currency identification method based on a mobile phone as an assistive aid for the visually impaired. The approach utilized the k-means clustering technique to recognize US banknote captured by blind users. The system extracted the SIFT features from the note, which are computationally demanding, thus increasing the recognition time. Also, a researcher implemented a recognition system based on a bionic eyeglass. This system extracted the Zernike moment and employed K-means and CNN for note identification. It was designed using a Samsung Galaxy S mobile phone, and five visually impaired subjects participated in the experiment. However, the method used a server/client architecture to transmit the image to a computer system for recognition, thus making the mobile phone act just as an input/output device with no processing capability [86].

Additionally, [108] developed an Indian note recognition system for mobile phones. The system engaged the visual BOW recognition method and used the GrabCut algorithm to separate the foreground from the background. It was tested on 2584 images and had an accuracy rate of 96.7%. However, the system was prone to failure based on image illumination and the inclusion of background in the captured banknote image. Finally, [32] designed a method for a currency recognition system. Their system extracted and compared the recognition rate of SIFT features from both grayscale and colored images. It was tested on 400 Jordanian currencies (coins and banknote). Still, the presented approach had illumination difficulty, was dependent on the capturing distance of the images, and was relatively slow when used on a mobile phone.

2.1.5 Banknote Counterfeit Detection

Over the years, several banknote counterfeit detection systems have been designed. [131] developed a banknote authentication system for Bangladeshi banknote; the system scans the micro-letter print on the note and feeds them into an optical character recognition engine that matches characters with stored prototypes. A method for the recognition of Sri Lankan was also recommended by [44], the system adopted a unique linear transformation function for noise

reduction, edge detection for fair representation of boundaries for the new and old damaged note and a three-layer back propagation Neural Network for Classification. Furthermore, [53] suggested a recognition system for worn-out banknote using dynamic spectrum classification based on the divergence-based kernel, the application of Support Vector Machine (SVM) and Relevance Vector Machine (RVM). Besides, [45] developed a paper currency recognition method using local binary pattern (LBP), the process was robust to noise and illumination change. The system was tested on RMB, although it was stated that it could be applied to banknote of other countries.

In the same view, [100] addressed a paper currency classification approach for authenticating the process used in printing different texts on currency note; nine features were extracted and classified using Support Vector Machines and Neural Nets. Furthermore, Linear Discriminate Analysis (LDA) was employed to authenticate the printing process of the discriminatory power of the selected features, and the result showed very high accuracy. Similarly, [5] presented software that can identify fake paper currencies from captured images. The system detects counterfeit paper money by extracting features like micro-printing, iridescent ink, optically variable ink, watermark, security thread and ultraviolet lines using optical character recognition technique. Also, the approach employed Contour analysis, Face Recognition, Speeded-up Robust Features (SURF) and Canny Edge & Hough transformation algorithm of OpenCV. Lastly, Ref. [51] investigated a framework that attempted to prove how simple changes in the design of banknote design and rendering systems can make the task of counterfeiting impossible.

2.2 Computer Aided Diagnosis

The field of medical pathology and radiology is tasked with providing absolute disease diagnoses to support and direct patient treatment and managerial decisions. Furthermore, the implementation of a metric to standardize the accuracy of diagnoses to ensure reproducibility that is essential for advancing precision in medical diagnosis has always been imperative. For over 100 years the main approach employed by pathologists for medical diagnosis was microscopes before advancements led to the invention of x-rays [35]. However, human interference based on visual analysis of radiology and microscopic images has been plagued by the lack of standardization, diagnostic blunders, and the substantial cognitive load required to evaluate the images manually. Consequently, in the last two decades, several researchers have seen the need to design

computational techniques that can assist in the analysis of medical pathology and radiology images [38].

CAD systems have focused primarily on machine learning methods using image analysis methods like normalization, segmentation, and feature extraction. After which classification models are constructed using standard machine learning methods including Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), K-Nearest Neighbor, and Decision Tree (DT) or deep learning based systems like CNN, RNN, and GANs with no discrete human-directed steps for segmentation, detection, and feature extraction. Instead, the deep learning algorithms take the images as the input only, its corresponding labels while training a very high-dimensional and complex set of model parameters with supervision come only from the image labels.

2.2.1 Convolutional Neural Network

The recent development in CNN has suggested the possibility of automatically uncovering features from the training images and exploiting the interaction within the deep structure of a neural network. Consequently, the issues of feature computing, selection, and integration can potentially be addressed by this new learning framework. As this method can directly uncover features from the training data, the effort of explicit elaboration on feature extraction can be significantly alleviated. Furthermore, the neuron-crafted features may compensate and even surpass the discriminative power of the conventional feature extraction techniques [28]. Additionally, feature interaction and hierarchy can be exploited jointly within the intrinsic deep architecture of a neural network thus significantly simplifying the feature selection process [112].

[99] investigated three existing CAD systems used in diagnosing colonic polyps, inflamed lymph nodes and sclerotic spine metastases on body CT. They employed 2D patches in three orthogonal directions, and up to 100 randomly rotated views. The CNN predictions were then aggregated to improve the system accuracy, leading to the sensitivity for lesion detection improved 13–34% for all three CAD systems after adapting the CNN's. Likewise, [10] also examined the benefit of deep learning methods for chest x-ray analysis; the process explored transfer learning to test the ability of CNN system trained on non-medical images in the detection of different types of diseases in chest radiographs. It presented promising results demonstrating the viability of distinguishing pathology in chest x-rays using deep learning methods based on non-medical learning.

Additionally, [8] focused on analyzing patterns of interstitial lung diseases from 2D patches of chest CT scans. They are one of three groups in this issue using a public dataset. They train CNN to classify patches of 32x32 pixels into one of 7 classes and report higher accuracy than three previously published methods that employ hand-crafted features for chest CT scans diagnosis of interstitial lung diseases. Furthermore, [31] detected cerebral microbleeds from susceptibility weighted MRI scans using 3D CNN's and replaced the candidate detection stage with a CNN thereby proposing a two-stage approach. They report improved results with their proposed 3D CNN compared to various classical, and 2D CNN techniques from the literature that the authors re-implemented and trained and tested on the same dataset.

While [107] introduced CAD system that utilizes a CNN combination from three previously designed pulmonary nodules detectors that extract 2-dimensional patches from nine different orientations of 3-Dimensional Chest CT scans, presenting a slight improvement in accuracy than the three previously developed nodule detectors. Furthermore, [109] developed a CNN method for identification and detection of nuclei in histopathological images. The system extracted small patches of histopathological images in combination with a fusion of overlapping patches had employed them as input data. The system did not just predict if the central pixel of the patch is a cell nucleus, it was also able to model the output as a high peak near the center of each nucleus and flat elsewhere. [125] also evaluated the computer-aided diagnosis of whole slide images containing metastatic breast cancer of sentinel lymph node biopsies of the classification task. A pathologist independently diagnosed the same images, and the knowledge was combined with our deep learning system’s predictions thus increasing the pathologist’s accuracy representing an increase in recognition rate while reducing the human error rate.

2.2.2 Generative Adversarial Network

Recently, deep learning has been introduced to the medical image analysis domain. GANs train a generative model by playing an adversarial game between a generator and an auxiliary discriminator, which classifies data samples versus generated ones. The generator employs noisy data to produce possible sample images to deceive the discriminator while the discriminator tends to improve its discriminative ability by learning. Recently several variations of GANs have been proposed to improve image generation and discrimination [40].

A researcher [22] developed an in-depth perceptual similarity metrics based on a class of loss functions; the method calculated the distances between image features extracted using deep neural networks to generate high-resolution images. The approach was applied to feature inversion and random natural image generation, it proved that the loss is superior to the typical loss in image space. It allows images to generate and include perceptually essential details even from very low-dimensional image representations. [76] also designed an unrolled generative adversarial network that aims to stabilize GANs by defining the generator objective by allowing training to be finetuned using the optimal discriminator in the generator's objective, the method stabilized the training of GANs with multiple recurrent generators, increases diversity and coverage of the data distribution of the generator. The Optimal discriminator is ideal but not feasible in practice in the generator's objective. The discriminator of GANs is often unstable and leads to poor solutions. This innovation allows training to be finetuned using the discriminators.

Furthermore, [14] introduced a boundary equilibrium generative adversarial network that employed a new equilibrium enforcing practice combined with a loss derived from the Wasserstein distance for training auto-encoder based GANs to balance the discriminator and generator. The technique uses a proposed GAN structure with a standard training method to provide a new approximate convergence measure, fast and stable training. Additionally, [74] formulated a Least squares generative adversarial network that implemented a least squares loss function for the discriminator with the view of minimizing the Pearson divergence, the method was more stable during the learning process and generated better quality than the ones created by regular GANs. The problem with this method is that during the learning process the loss function may lead to the vanishing gradients. In the same view, [117] outlined a novel generalized framework for adversarial adaptation. They employed an unexplored instance of a general structure which combines discriminative modeling, free weight sharing, and a GAN loss, which is called Adversarial Discriminative Domain Adaptation (ADDA). It shows that ADDA is more effective yet considerably simpler than competing domain-adversarial methods and demonstrates the potential of the approach by exceeding different unsupervised adaptation results on standard cross-domain digit classification tasks and a new more difficult cross-modality object classification task.

Similarly, [65] developed a Face Generative Adversarial Network that can stabilize the loss fluctuation while improving the capabilities of generator and discriminator. The performance of the method in comparison with existing Deep Convolution Generative Adversarial Network using

the Celeba dataset. Lastly, [130] suggested an Autoregressive Generative Adversarial Network that model rather than relying on binary classification of samples into data/generated classes using the latent distribution of data in an autoregressive model, However, the model can be further enhanced by more advanced latent space modeling, such as 2D LSTM, hierarchical modules, and bidirectional modules.

2.2.3 Single Image Super-Resolution Generation

Several researchers have addressed the image super-resolution problem; prior algorithms utilized filter techniques like bilinear, bicubic, and Lanczos filtering. However, these approaches were only able to generate smooth output without recovering any high-frequency information. [6] suggested edge features that generate high-resolution edge map using a rectangular center-on-surround-off filter before performing piecewise linear interpolation of the zero crossings in the output of the filter. [64] exploited the interpolation constraint using the geometric duality between the high-resolution covariance and its corresponding low-resolution covariance. The accuracy was limited due to the oversimplification of the SR problem, but the approach was computationally efficient. [135] developed a technique to enhance visual effects by using a multi-scale dictionary to capture and upscale samples of similar image patches at different scales. [116] precomputed the corresponding embedding matrix and anchored the neighborhood embedding of a low-resolution patch to the nearest atom in the dictionary while enhancing the approach by using the training data in the testing procedure.

[39] introduced a technique that took into consideration self-similarity and employed example-based SR and classic multi-image SR utilizing patch redundancies across image scales. [52] employed a geometric variation to expand the internal patch search while explicitly localizing planes in the scene, and applying the detected perspective geometry in guiding the patch search process. [126] utilized a feed-forward network based on the learned iterative shrinkage and thresholding algorithm integrated with a sparse representation prior. [60] applied a deep Convolutional Neural Network (CNN) based on recursive structures and skip connections that has 20 layers in addition to a small filter with a high learning rate and an adjustable gradient-clipping. The approach used a limited model parameter to identify that for long-range pixel dependencies using a highly performant architecture. [43] proposed a method that reconstructs realistic texture

detail while avoiding edge artifacts using a convolutional sparse coding approach that improves consistency by processing the whole image rather than using overlapping patches.

Recently, CNN's have widely been implemented for image SR, [30] developed a trained a three-layer deep fully convolutional network that used a bicubic interpolation to upscale an input image end-to-end to achieve state-of-the-art SR performance. It presented a super-resolution convolutional neural network. It is a novel deep learning approach for super-resolution of a single image by restructuring a deep convolutional neural network from the traditional sparse-coding-based SR techniques. The method learns an end-to-end mapping between the low- and high-resolution image and jointly optimizes all layers in the network, unlike conventional systems that analyze each component separately. The technique is simple, robust, and could be applied to other low-level vision problems achieving state-of-the-art restoration quality, and speed for practical on-line usage. It proves that enabling the network to learn the upscaling filters directly can further increase performance both in accuracy and speed. Meanwhile [55] and [16] employed loss function closer to perceptual similarity to regain visually more finite HR images.

[73] introduced a PixelGAN generative autoencoder that combines a generative PixelCNN conditioned on a latent code with a generative adversarial network (GAN) inference network that can impose arbitrary prior to the latent code distribution. The technique proved that different priors result in different decompositions of information between the latent code and the autoregressive decoder thus showing that the PixelGAN autoencoder with a categorical prior can be directly employed in semi-supervised sceneries with competitive results on MNIST[27], SVHN and NORB dataset.

[25] proposed a new deep network architecture. It is a pixel recursive super-resolution model that extends the PixelCNN. The technique is a fully probabilistic approach that tackles super-resolution with small inputs, demonstrating that images can be enlarged with improved resolution. The model produces a diversity of plausible high-resolution images at large magnification factors and shows how high-resolution images sampled from this network can fool a naive human observer a significant fraction of the time. [62] applied a deep residual network for image super-resolution (SR); the SRGAN technique could infer photo-realistic original images four times upscaling factors. They proposed a perceptual loss function which comprises a content loss and an adversarial loss. The content loss is driven by perceptual similarity rather than the similarity in pixel space; while the adversarial loss uses a discriminator network that is trained to differentiate

between the super-resolved images and original photo-realistic images to push our solution to the natural image manifold. The method could recover photo-realistic textures from heavily downsampled images on public benchmarks with more photo-realistic than reconstructions obtained with state-of-the-art reference methods.

Lastly, [97] proposed a cascade training approach to deep learning to improve the accuracy of the neural networks using SRCNN while gradually increasing the number of network layers using the cascade trimming approach. The method gradually reduces the layer of the network layer by layer without significant loss on its discriminative ability.

Chapter 3

3 Banknote Recognition System

3.1 Multi-Class SVM Based Gradient Feature for Banknote Recognition

This chapter will first present a banknote recognition system based on a PCA-HOG feature. Furthermore, an efficient error correcting output code technique based on Multi-Class SVM for identification of note denomination is presented. The method had three main phases: background subtraction, feature extraction, and classification.

3.1.1 Background Subtraction

During image acquisition, the irrelevant background image was captured. To efficiently identify the banknote, this unwanted data must be removed [69]. Figure 1 presents samples of (a) Full note, (b) Partial note and (c) Old note of Nigerian Naira, US Dollar, Canadian Dollar, and European Euro utilized during this research. In this work, background subtraction was achieved using the following steps:

Step 1: The colored image was converted to grayscale by excluding the hue and saturation while preserving the luminance. Then, the intensity image was converted to a double precision format to aid binary conversion.

Step 2: Otsu threshold [137] was used to turn the image into a binary image by minimizing the intra-class variance of the black and white pixels. All holes in the binary image were located and filled using the imfill method [1].

Step 3: The connected components of all sets of objects in the image was computed. Then the exterior boundaries of the objects, including edges of holes inside the objects were traced. The largest object was located and chosen as the banknote.

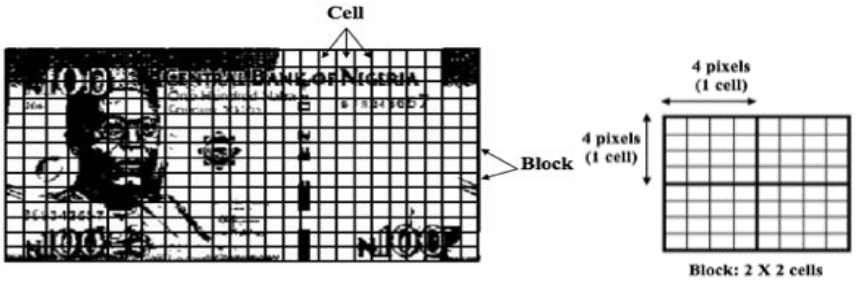
a) b) c)

Figure 1. Sample (a) full note, (b) partial note, and (c) old note

Step 4: The area and centroid of the note were calculated. All non-Naira pixels were set to zero, and the bounding box of the note was used to extract the banknote from the image. Finally, the bill was normalized to a fixed size of 154 by 302. The binary image was converted to grayscale since the proposed method conducts its feature extraction on grayscale images.

3.1.2 Feature Extraction

Histogram of Gradient is a well-known feature extraction technique that is robust to illumination, positioning, and change in shape [57]. The method was introduced as a pattern recognition method in areas like vehicle recognition, pedestrian recognition (standing humans, walking human and people riding a bicycle), and image word spotting. Nowadays, it has been used for object classification [56] and if properly utilized, can be applied to banknote recognition.



(a) (b)

Figure 2. (a) Cell, and (b) block module of the HOG method

Before extracting the HOG, the system divided the pre-processed image into small spatial sections called cells; computed the gradient of each cell, and its results for the directions θ ($00 \leq \theta$

≤ 1800) separated by an interval of 20 degrees. Additionally, the cells were grouped into a predefined dimensional vector called a block by the addition of the magnitude m and the block was generated throughout the image based on the cell size [75]. Figure 2 depicts the cell and block components of the HOG technique. The cell creation involved partitioning the image into groups made up of 16 pixels, and the blocks derived from combining four cells. Furthermore, the process of extracting the HOG feature for each pixel was usually divided into three stages: calculation of gradient, generation of the histogram, and normalization of the histogram. This research inserts an additional step that extracts the principal component of the HOG features, and this is shown in Figure 3.

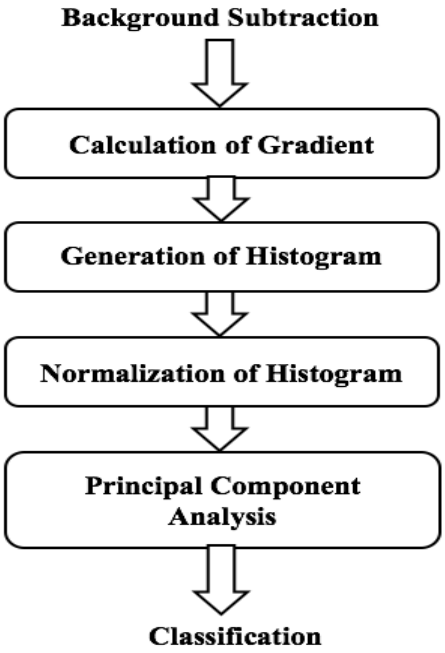


Figure 3. Stages of PCA-HOG extraction

Calculation of Gradient

To calculate the gradient, the First Order Differential coefficient was computed using equation (1) [4].

$$\begin{cases} f_x(x,y) = f(x+1,y) - f(x-1,y), \\ f_y(x,y) = f(x,y+1) - f(x,y-1), \end{cases} \quad 1$$

The First order coefficient was then applied to calculate the magnitude m (*x, y*) and direction θ (*x, y)* of the gradients using Equations 2 and 3 respectively [127].

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad 2$$

$$\theta(x,y) = arctan\ \frac{f_x(x,y)}{f_y(x,y)} \quad 3$$

Generation of Histogram

After calculating the image gradients, the magnitude was employed to compute the histogram of each cell. After that, the direction function aided in reducing the number of details present in the gradient of each cell [56].

Normalization of Histogram

Histogram normalization was achieved by combining the histograms of all cells into a predefined block size using Equation 4.

$$\upsilon = \frac{\mathbb{V}_k}{\sqrt{||\mathbb{V}_k||^2 + 1}} \quad 4$$

where $\mathbb{V}_k$ is a vector derived from the combination of the histograms of all cells grouped into blocks [63].

Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised scheme that utilizes a linear transformation method in reducing the size of vectors without affecting the variation in the data. This method calculates the eigenvalues and eigenvectors of the data by selecting the eigenvector with the largest Eigen Values as a principal component of the dataset and uses it in forming a new dataset based on the vectors that map the data to the new feature space [122].

3.1.3 Classification

Multi-Class SVM has been utilized in different fields such as text recognition, face recognition, and traffic sign recognition. Figure 4 presents the structure of the Multi-Class SVM. The system maps the input features using a constructed linear decision surface in a higher dimensional feature space [81]. It resolves all negative samples and uses the SVM decision functions in classifying the new data points. Multi-Class SVM has two application approaches One Versus All (OVA) method and One Versus One (OVO) method. The OVO technique teaches the system to differentiate positive and negative class from the training set. On the other hand, the OVA technique trains the

classifier one class at a time using the positive class and groups all other samples as belonging to the negative class [50].

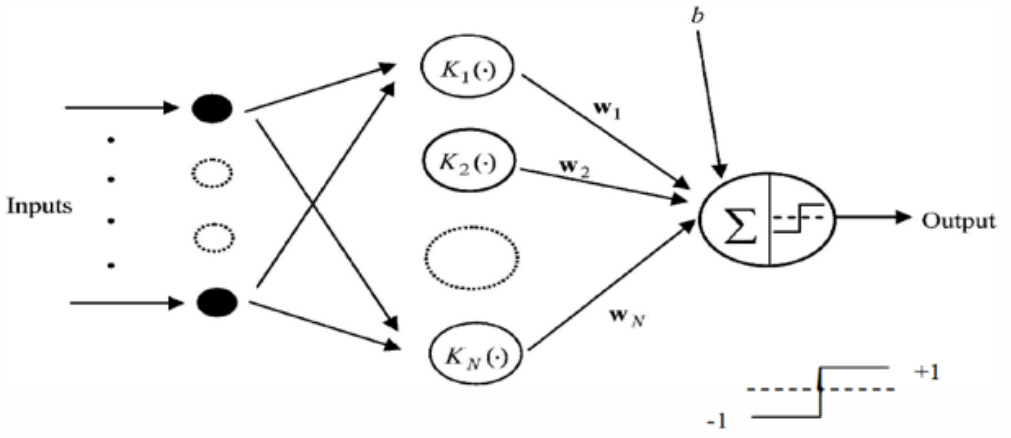


Figure 4. General structure of the MCSVM

For this research, the OVO approach was applied. The method trained one class at a time, and for each class, it divided the training set into two categories: positive and negative classes while ignoring the remaining classes. The system resolved all negative samples and used the SVM decision functions in classifying the new data points [9]. In conclusion, the ECOC based Multi-Class SVM reduces noise in classification data caused by the complexities of class divisions, training sets, and base classifier overfitting that can lead to misclassification [49].

3.1.4 Data Collection

In this research, the banknote image was captured using a Canon T2i digital camera with a resolution of 200 dpi and 24 bit picture scan mode. Since there is no publicly available dataset of Nigerian Naira (₦5, ₦10, ₦20, ₦50, ₦100, ₦200, ₦500, ₦1000), US Dollar ($1, $5, $20, $50, $100), Canadian Dollar ($5, $10, $20, $50, $100), and Euro (€5, €10, €20, €50, €100, €200, €500), this research created a database called Naira_00N, USD_00D, CAD_00D, and EUR_00E. The statistical distribution of the Naira, US Dollar, Canadian Dollar, and Euro dataset comprised of 1800, 1000, 1000 and 1400 samples respectively. Furthermore, the figures presented the number of new, old, full and partial note utilized in this research. Finally, the dataset was made up of images captured in four directions for each banknote: the front side (upward and downward) and the backside (upward and downward).

3.1.5 Experimental Setup

The hardware utilized comprised a Core i3, 2.4 GHz processor with 8 GB Ram. The operating system used was Windows 8.0 64 bit. Image processing techniques and other related MATLAB toolboxes were utilized in the development of the system. The research utilized Cross-Validation for data analysis. It is a statistical technique used in evaluating predictive models by segmenting data into training and validation sets. The approach randomly divided the dataset into K-fold equal data sizes and crossed over each K-dataset partition in successive rounds such that each data point has a chance of being in the validation process. A 10-fold Cross-Validation approach was employed in the classification of the Naira, USD, CAD, and Euro databases.

3.1.6 Result and Evaluation

Table 1. Recognition result for CAD dataset (PCA-HOG+MCSVM)

Classifier	**Recognition Rate**			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Naïve Bayes Multinomial	0.953	0.944	0.945	0.944
Random Forest	0.958	0.956	0.956	0.956
NNge	0.919	0.889	0.895	0.889
Hyper pipes	0.918	0.911	0.91	0.911
IB1	0.94	0.939	0.939	0.939
SMOSVM	0.953	0.95	0.95	0.95
MCSVM	**0.963**	**0.96**	**0.961**	**0.96**

Table 2. Recognition result of USD dataset (PCA-HOG+MCSVM)

Classifier	**Recognition Rate**			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Naïve Bayes Multinomial	0.927	0.92	0.92	0.92
Random Forest	0.927	0.92	0.92	0.92
NNge	0.916	0.89	0.89	0.89
Hyperpipes	0.895	0.978	0.764	0.978
IB1	0.927	0.92	0.92	0.92
SMOSVM	0.932	0.93	0.93	0.93
MCSVM	**0.971**	**0.97**	**0.97**	**0.97**

In this section, we appraise the performance of the proposed technique. Tables 1, 2, 3 and 4 present the recognition rates of CAD, USD, Euro, and Naira note respectively. It shows the

precision, recall, f-measure, and accuracy for each classifier utilized in this research. A note was classified using the MCSVM, and for experimental evaluation, six other classifiers were utilized in classifying the note, namely: Naïve Bayes Multinomial, Random Forest, NNge, Hyper pipes, IB1, and SMOSVM. The results show that the suggested MCSVM classifier had the best recognition rate in comparison with existing methods used in banknote classification.

Table 3. Recognition result of Euro dataset (PCA-HOG+MCSVM)

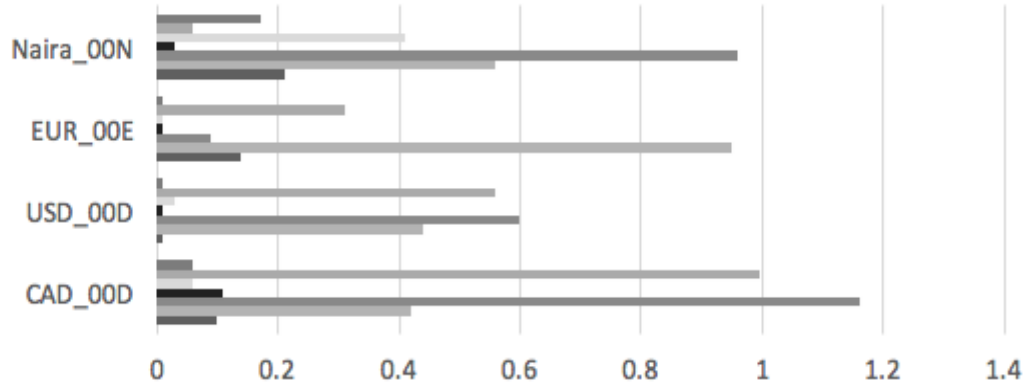
Classifier	**Recognition Rate**			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Naïve Bayes Multinomial	0.957	0.95	0.949	0.95
Random Forest	0.891	0.85	0.834	0.85
NNge	0.982	0.98	0.98	0.98
Hyperpipes	0.888	0.77	0.77	0.77
IB1	0.901	0.89	0.886	0.89
SMOSVM	0.956	0.95	0.949	0.95
MCSVM	**0.986**	**0.986**	**0.986**	**0.986**

Table 4. Recognition result of Naira dataset (PCA-HOG+MCSVM)

Classifier	**Recognition Rate**			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Naïve Bayes Multinomial	0.986	0.986	0.986	0.986
Random Forest	0.986	0.986	0.986	0.986
NNge	0.972	0.971	0.972	0.971
Hyperpipes	0.96	0.957	0.958	0.957
IB1	0.986	0.986	0.986	0.986
SMOSVM	0.979	0.979	0.979	0.979
MCSVM	**0.99**	**0.99**	**0.99**	**0.99**

Figure 5 presents the average processing time for each dataset with each classifier as each image varied because of the differences in image size, quality, and pixel density. The Euro had the shortest processing time when classified using the IB1, SMOSVM, NNge, and Hyper pipes, while the USD presented the shortest recognition time with the Naïve Bayes Multinomial. Furthermore, the CAD also had the shortest processing time with the Random Forest, and finally, the Naira note had the shortest time with the proposed classification method MCSVM. From the timing, we concluded that the banknote dimensions caused a variation in the image processing time. The feature extraction time also dispersed as a function of the number of pixels, cells, and blocks

utilized in each dataset. However, the difference between the validation time for each classifier was small regardless of the dataset. In conclusion, although all other classifiers have remarkable recognition rates and computation time, the suggested MCSVM classification method presented the best results.



	CAD_00D	USD_00D	EUR_00E	Naira_00N
MCSVM	0.06	0.01	0.01	0.17
SMOSVM	0.995	0.56	0.31	0.058
IB1	0.06	0.03	0.01	0.41
Hyper Pipes	0.11	0.01	0.01	0.03
Nnge	1.16	0.6	0.09	0.96
Random Forest	0.42	0.44	0.95	0.56
Naïve Bayes Multinomial	0.1	0.01	0.14	0.21

Figure 5. Processing time(sec) of dataset 1 – 4 (PCA-HOG+MCSVM)

3.2 Modified HOG Descriptor-Based Banknote Recognition System

In this chapter, we implemented a procedure for enhancing and cleaning up banknote images. It involves converting to grayscale then employing Gamma Correction and Histogram equalization to improve the image luminance and contrast of the intensity image. Furthermore, Weiner filter was applied to reduce the amount of noise. After that, the boundaries of the banknote are computed using Blob detection, cropped out while Gaussian pyramid is utilized in standardizing the note. We also created a new feature set by extracting the Histogram of Gradient (HOG) Descriptors from image patches generated from the vertices of SURF interest points. After the HOG features have been obtained, the vector size may differ for each image in addition to being computationally

demanding during processing. To minimize its effect on recognition/detection rate while maximizing the reduction of Feature dimensionality. A combination of dimensionality reduction techniques based on the descriptor correlation and variance is then applied to reduce the Vector scale. A descriptor filter analysis using Pearson's Product Moment Coefficient ρ and variance σ^2 is employed to determine the best set of features for classification.

The approach answers two fundamental questions by determining the relevance of the descriptor to the feature set and its redundancy regarding other associated descriptors. Finally, we applied our technique to several datasets using OVO-SVM and Putative point matching to investigate the efficiency of the descriptor showing the effectiveness of Modified HOG/LVHCF descriptor for banknote identification, counterfeit detection, and multiple note recognition. Figure 6 shows the phase of the proposed Modified Hog-LVHCF based paper money recognition system.

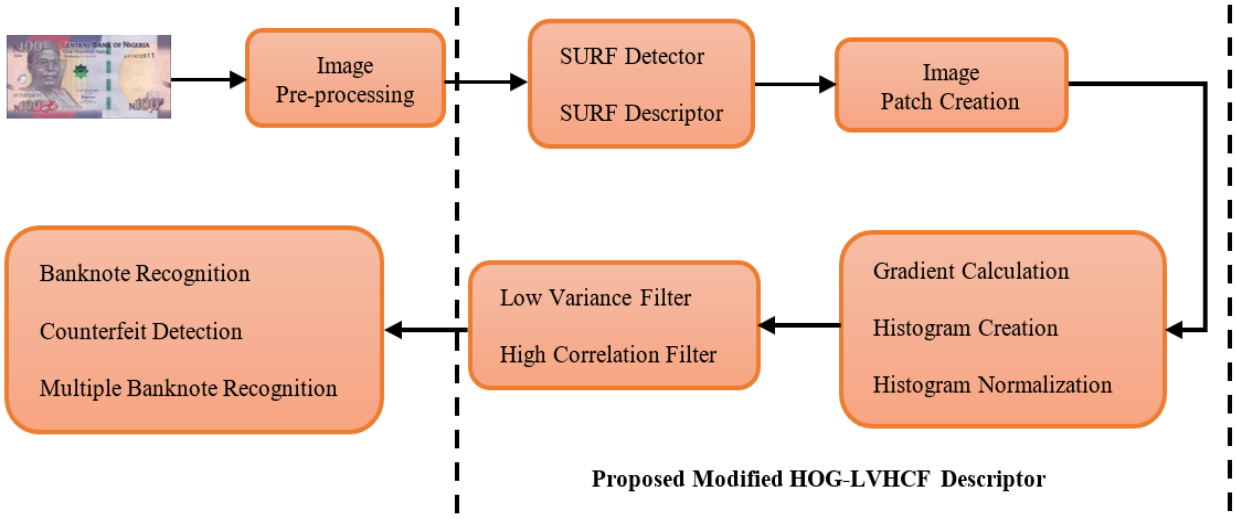


Figure 6. Proposed modified HOG-LVHCF based banknote recognition system

3.2.1 Image Pre-processing

During image acquisition, the irrelevant background image is also captured. This unwanted information must be removed to identify and recognize the note. Furthermore, digitized banknote images sometimes have discontinuous particles, noise, illumination, positioning, and unanticipated boundaries. Also, when matching the image. If the banknote is of varied sizes it could affect the recognition rate. In this research, image pre-processing plays a fundamental step as it involves enhancing and cleaning up the images for efficient feature extraction. Thus, in this research, a new

combination of techniques for image pre-processing is presented, and a flowchart of the procedure is shown in Figure 7.

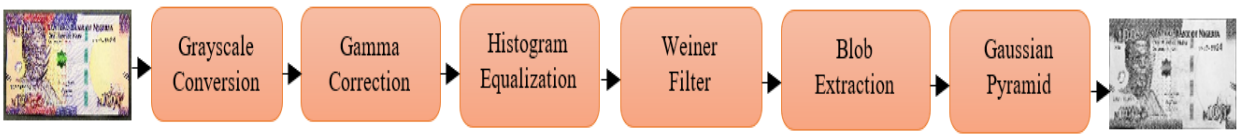


Figure 7. Block diagram of the procedures undergone in the image pre-processing

Step 1: The image was first converted to grayscale by calculating the weighted sum of the three channels of the RGB image and representing the result as a matrix whose fragments are assigned values based on how dark or bright the color of the pixel is at the corresponding position. The value allocated to the pixel during grayscale conversion ranges from 0 to 255, representing the brightness of the image.

Step 2: Gamma Correction was applied to improve the image luminance by enhancing the quality of the image in dark or blurry areas while compressing the intensity of bright regions.

Step 3: Histogram equalization was then employed in enhancing the contrast of the intensity image. The approach redistributes the image histogram to ensure that the brightness value is constant across the image.

Step 4: Weiner filter was applied to reduce the amount of noise present in a signal comparing with an estimate of the desired noiseless signal.

Step 5: Background subtraction was achieved by first finding all blobs in the image using Blob detection, select the boundary coordinates of the blob containing the banknote and cropping out the note using the detected blob edges.

Step 6: Banknote denominations' sizes vary, and the note can also change due to the acquisition manner thus image reduction and standardization were achieved using the Gaussian pyramid.

3.2.2 Feature Extraction (Proposed Modified HOG Descriptor)

Feature extraction is a crucial step in image processing, the selection of features is vital in determining the recognition of an object, and an ideal feature set should be stable, invariant to rotation, scale, and illumination changes. Several feature extraction methods have been proposed over the years. In this research, several useful descriptors and feature reduction approaches were

investigated and compared to the proposed Modified HOG descriptor and Low Variance High Correlation Filter. In this study, six different descriptors were employed: Scale Invariant Feature Transform, speeded up robust features, gradient location and orientation histogram, Histogram of Gradient, shape context, and the proposed Modified HOG Descriptor. To reduce the processing time, several feature reduction methods were also implemented: Principal Component Analysis, Bag of Words, High Correlation filter, Low Variance Filter, while comparing to the developed Low Variance and High Correlation Filter Algorithm. The proposed Modified HOG based descriptor technique is split into four phases: SURF point detection, SURF descriptor extraction, image patches creation from SURF points, and HOG feature computation.

SURF Point Detection

The SURF interest point detection is separated into four phases: integral image generation, calculation of the matrix, construction of the scale space and localization of the feature point.

Integral Image Generation: The Integral image generation was introduced to reduce the computational complexity of the square like convolution filters used in the future, it is mapped from the original image. The integral image is then computed by the sum of all pixels in the input image *I* within a given rectangle sector which can be deduced by Equation 5.

$$\mathrm{I}_{\Sigma}(\mathrm{x},\mathrm{y}) = \sum_{\mathrm{i}=0}^{\mathrm{x}} \sum_{\mathrm{j}=0}^{\mathrm{y}} \mathrm{I}(\mathrm{i},\mathrm{j}) \qquad 5$$

After the integral image has been calculated, three additions are carried out to compute the summation of the intensities over any four-sided vertical area consequently the computational time is not dependent on the image size using Equation 6.

$$\sum = \mathrm{I}_{\Sigma}(\mathrm{A}) + \mathrm{I}_{\Sigma}(\mathrm{D}) - \left(\mathrm{I}_{\Sigma}(\mathrm{C}) + \mathrm{I}_{\Sigma}(\mathrm{B})\right) \qquad 6$$

Computation of the Hessian Matrix: The box filters are used as an approximation of the second-order Gaussian derivatives, this is computed on the integral image calculated earlier rather than the actual image hence reducing the processing time. A 9 x 9 square-like filter was employed in the computation of the Gaussian denoted by with a scale of 1.2 which led to the creation of a blob response map over different scales in the image. A set of simple weights *w* were applied to the square-like sector to reduce the computational overhead.

Construction of the scale space: The SURF approach is scale invariance, it applies the same filter at different sizes to the image at the same speed in parallel. The length of the dark lobe can only be increased by an even number of pixels to guarantee the presence of a central pixel (top). The scale space is categorized as a series of octaves; an octave is a set of filter response maps gotten by convolving each input image with a filter in different sizes. For an octave layer with a size N * N, the scale σ, the convolution of each layer in the image is computed on the response map using equation 7.

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (7)$$

Localization of the interest point: The final step in the interest point detection is localizing the image over the different scales, a threshold is set for removing weak responses in the map. Several thresholds are checked until a middle ground is reached that takes into consideration the number of interest points and the strength of the detected interest point. Furthermore, a 3-dimensional scale space is applied to the non-maximum suppression, and only points bigger than all the neighbors is selected as the given interest points. Finally, scale space interpolation is applied to obtain the best interest points [17].

SURF Descriptor Extraction

SURF Descriptor computation is divided into three phases:

1) Assignment of the interest point orientation

2) Computation of the descriptor using Haar Wavelet transform

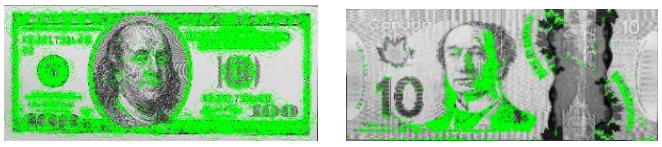
3) Feature matching using Fast Indexing [85].

Assignment of the interest point orientation: Assignment of Interest points orientation is needed to attain rotation invariance, so Haar Wavelet transform is used at each interest point (x, y) to determine the direction of a circular neighborhood in addition to its scale. After computing the wavelet transform, the feature point is weighted using a Gaussian filter, the dominant orientation is decided by calculating the summation of all the horizontal and vertical responses within the sliding windows; the most extended vector is then selected as the local orientation vector for the feature point.

Computation of the descriptor using Haar Wavelet transforms: The previous step used a box area centered around the point of interest in determining the orientation. In this phase, each sector is

further divided into a 4x4 square like sector, to improve the reduce the geometric distortions, and localization inaccuracies that could occur. The interest point is weighted with a Gaussian filter then the Haar wavelet transform is calculated horizontally and vertically for each 5x5 paced interest point. Furthermore, the computed wavelet transform is then summed for each sector and used as the first level entry in forming the feature vector set. Moreover, to capture the change in intensity after each interest point, the absolute value of the responses is computed to create a 4-dimensional descriptor vector for each sub-sector. Finally, the sectors which are made up of a 4x4 sub-sector are concatenated, resulting in a descriptor with a length of 64 dimensions.

Feature matching using Fast Indexing: This phase presents an additional functionality to the matching stage, it uses a fast indexing technique based on the previous computed Laplacian in the Haar wavelet transform stage. Figure 8 presents the SURF point descriptors of sample USD and CAD paper money.



a) b)

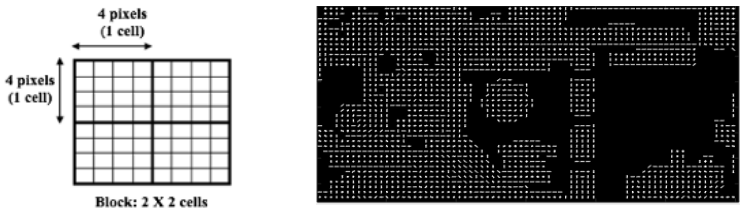
Figure 8. Detected SURF descriptors of a) the US, and b) Canadian dollar

Creation of Image Patches

An image patch is designed from polygons defined by the coordinates of its vertices; it consists of one or extra polygons that might or might not be connected. In this research, the 43 x 43 patch is formed by specifying the coordinates of each unique vertex of the SURF square-like descriptors and a matrix that defines how to connect these vertices to create the faces. This technique is preferred for SURF points as its patches multifaceted, so vertices shared by more than one face need be defined only once thereby requiring fewer data to design the patches.

HOG Descriptor Computation

Histogram of Gradient (HOG) feature is a viable feature extraction technique with the ability to select feature points irrespective of the banknote image view. It is based on the histogram of pixel gradients neighbors for image sections or blocks thus it is hardly affected by changes in lighting conditions or image geometry. The HOG feature extraction technique can be divided into three levels: calculation of Gradient for each pixel, computing its Histogram and by combining the histogram of all cells in a block to normalizing the histogram [124].



a) b)

Figure 9. (a) Cell/block module, and (b) sample HOG on Naira note

Figure 9 shows the terminology of the HOG feature extraction. It depicts the gradient dispersion on the different positioning of a banknote image and is conducted to obtain the dimension and texture feature of the banknote. The algorithm partitions all note image into small evenly-dimensioned sections called cells comprising of several pixels which individually have a weight computed by getting the gradient magnitude and histogram. It shows the dispersion of the pixel taking into cognizance the orientation of the gradient. The cell histogram is then used as a vector that is merged to get the HOG feature of the overall banknote image. Finally, Table 5 shows the HOG method specification that provided the best result in this research with the Cell size of 16, four cells per block, 1 to 1 block overlap, nine bins, and 18x1 bin centers. The number of HOG feature is different for each dataset because it is chosen based on the feature size that presented the best recognition and detection rates.

Table 5. Specification of the HOG descriptor

Specification	Naira	USD	CAD	Euro
SURF Descriptor	124864	196240	209002	205855
Cell Size	16	16	16	16
Number of cells per block	4	4	4	4
Block Overlap	1,1	1,1	1,1	1,1
Number of bins	9	9	9	9
Number of HOG extracted	38102	56109	76464	70136
Bin centers	18 x 1	18 x 1	18 x 1	18 x 1

3.2.3 Feature Reduction (Proposed LVHCF)

After the HOG features have been extracted, the vector size is computationally demanding. In order to unify and reduce the feature dimension; the techniques aim to minimize its effect on recognition/detection rate while maximizing the reduction of Feature dimensionality. The experiment conducts a descriptor filter analysis using Pearson's Product Moment Coefficient ρ and

variance σ^2 to determine the best set of descriptors for classification. The technique answers two fundamental questions: determining which descriptor is relevant or not to the feature set and if the descriptor is redundant or not concerning other related descriptors.

$$\rho = \frac{m(\sum ij) - (\sum i)(\sum j)}{\sqrt{[m\sum i^2 - (\sum i)^2][m\sum j - (\sum j)^2]}} \quad 8$$

$$\sigma^2 = \frac{\sum(x - u^2)}{m} \text{ where } \mu \text{ is the mean} \quad 9$$

The first question can be answered by just defining a threshold ρ value, but for the second question, it is more complicated as it involves analysis of pairwise correlations between all descriptors. To answer both questions before filtering the descriptors, we put forward an approach called a Low Variance and High Correlation Filter (LVHCF). In this experiment, it is assumed that if two descriptors are found to be redundant, one of them needs to be removed. By removing the one that is less relevant to the descriptor-set concept, it keeps more information to predict the class while reducing redundancy in the data.

Table 6. Feature reduction on dataset

Technique	Naira	USD	CAD	Euro
PCA	680	740	690	755
BOW	600	650	650	670
LVF	1905	2805	2293	3507
HCF	1143	1685	2294	2104
LVHCF	581	661	764	701

As presented in Figure 10, given a dataset with *S* feature set and a descriptor set *ds*, the algorithm chooses a list of predominant descriptors D_{best} for a given ds. It comprises two phases, the first phase (line 2-17) computes the Pearson correlation moment coefficient p for all pairs of descriptors in the ds using equation 8, extracts relevant descriptors into $S1^i_{list}$ based on the predefined threshold *t1*. The $S1^i_{list}$ is further processed to remove additional correlated pairs and only keeps predominantly uncorrelated *Dq* and *Dp* among all the earlier selected descriptors in the $S1^i_{list}$ (line 13). After one round of filtering all *Dq* pairs based on *Dp*, the algorithm will take the currently remaining feature right next to *Dp* as the new reference (line 18) to repeat the filtering process. The algorithm stops until there is no more feature to be removed from $S1^i_{list}$. In the second phase (line 19 – 38), the variance σ^2 of each descriptor in the $S1^i_{list}$ is computed using equation 9.

Input: S (D_1, D_2, …, D_m, ds)// training dataset

t1 = 0.287 // a predefined threshold (ρ)

t2 = 0.045 // a predefined threshold (σ^2)

output: S_{best} // an optimal subset

```
1  Begin
2    For i = 1 to M
3      For j = 1 to M
4    Do begin
5      Compute ρ_{i, ds} for all D_i and D_j
6      if (ρ_{i, ds} >= t1)
7        Append D_i to S1^i_list;
8    End;
9    D_p = getFirstDescriptor (S1^i_list);
10   D_q = getNextDescriptor (S1^i_list);
11     Compute ρ for all D_p and D_q
12     If (ρ >= t1)
13       remove D_q from S1^i_list;
14         D_q = getNextDescriptor (S1^i_list);
15     Else D_q = getNextDescriptor (S1^i_list);
16     End until (D_q == NULL);
17     D_p = getNextDescriptor (S1^i_list, D_p);
18   End until (D_p == NULL);
19   For i = 1 to M
20   Do begin
21     Compute σ^2_{i, ds} for S1^i_list;
22     if (σ^2_{i, ds} <= t2)
23       append D_i to S2^i_list;
24   End;
25   Order S2^i_list in ascending σ^2_{i, ds} value;
26   D_p = getFirstDescriptor (S2^i_list)
27   Do begin
28     D_q = getNextDescriptor (S2^i_list, D_p);
29     if D_q <> NULL)
30       Do begin
31         D'_q = D_q;
32           if (σ^2_{p, q} <= σ^2_{q, ds})
33             remove D_q from S2^i_list;
34             D_q = getNextDescriptor (S2^i_list, D'_q);
35           Else D_q = getNextDescriptor (S2^i_list, D_q);
36       End until (D_p == NULL);
37     D_p = getNextDescriptor (S2^i_list, D_p);
38   End until (D_p == NULL);
39 S_best = S2^i_list;
40 End;
```

Figure 10. Low variance and high correlation filter (LVHCF) algorithm

Furthermore, the descriptors are extracted into a new list $S2^{i}_{list}$ based on the predefined threshold *t2* and ordered in ascending order according to their σ^2 values. The list is further processed to remove redundant descriptors and only keeps predominant ones among all the selected relevant descriptors. A descriptor *Dp* that has already been determined to be dominant can always be used to filter out other features that are ranked higher than *Dp* and have *Dp* as one of its redundant σ^2 peers. The iteration starts from the first descriptor in $S2^{i}_{list}$ (line 26) and continues as follows. For all the remaining descriptors, if *Dp* happens to be a redundant peer to a feature *Dq*, *Dq* will be removed from $S2^{i}_{list}$. After one round of filtering descriptors based on Dp, the algorithm will take the currently remaining descriptors right next to *Dp* as the new reference (line 37) to repeat the filtering process. The algorithm stops when there is no more feature to be removed from $S2^{i}_{list}$. The descriptors matching the variance remaining in $S2^{i}_{list}$ is then used as the final dataset B_{est}. Table 6 presents the feature reduction on the four datasets as shown on Page 36.

3.2.4 Classification

The classification stage is the decision segment of a recognition system, and it uses the features extracted in the previous phase. It involves identifying each banknote while assigning it the correct denomination class. Furthermore, the principal function of a pattern recognition system is to yield decisions concerning the class membership of the models with which it is confronted. Support Vector Machine (SVM) is a classification approach used for separating hyperplanes. It was designed in the 1990s for building nonlinear separating function, and real-valued estimation [120]. The Multi-Class SVM-based system uses a binary classification approach for the one versus one technique. It designs a classifier for each class by creating a hyperplane between the classes. Furthermore, the system assigns a class to a test sample based on a distance margin for the classes. Each dataset is split into two groups; the training, and test groups. The system used a randomized approach for data selection, and the class of the test data was predicted based on the established class of the training mechanism. The Multi-Class SVM has two application method approaches; One Versus All (OVA) method and One Versus One (OVO) method. OVA differentiates one class from the remaining classes, and the OVO distinguishes each class from one class at a time while ignoring the remaining class [94].

For this research, the system used the OVO technique in combination with error correcting output codes. The error correcting output code divides a Multi-Class scheme into a double class

system that constructs a classifier for each binary division. OVO technique compares all classes against each other and builds a binary classifier to distinguish each pair of class while ignoring the remaining classes. When classifying an unknown image to a class, the classifier conducts a vote, and the class with the highest majority is selected as the best choice to assign the banknote [41]. Further analysis was conducted using K-Nearest Neighbor (KNN), Weighted K-Nearest Neighbor (WKNN), Random Forest (RF) Decision Tree, and multilayer perceptron (MLP) Neural network. From there, it is hard to decide which classification method is the best as research has only been able to prove that a classifier could have higher recognition and detection rates than others based on a specific problem.

3.2.5 Data Collection

For the dataset employed for denomination identification, the obverse and reverse of the banknote were captured using a Canon T2i digital camera with 18 Megapixels. Banknote suffering from scratches, wrinkle, stains, wear and tear which represent the image quality of the paper monies in the real-world were captured for this project. The statistical distribution of the dataset was Nigerian Naira, US Dollar, Canadian Dollar, and Euro datasets. It comprised 1790, 1120, 1000, and 1400 genuine banknote and 1320, 823, 733, and 1020 fake banknote respectively. Each dataset included images with multiple notes, clutters, overlapping and partial paper money. Furthermore, the dataset was made up of images captured in four directions for each banknote: the front (upward and downward) and back sides (upward and downward). The counterfeit currency dataset consisted of Naira note, Canadain Dollar note, Euro bills, and US Dollar covering a wide variety of conditions, such as occlusion, rotation, changes in scaling, illumination, and viewpoints. Lastly, the multiple banknote sets contained 460 Naira images, 372 US dollar samples, 440 Canadian dollar images, and 365 Euro Samples ranging from images with one to five banknotes.

3.2.6 Experimental Setup

This system was implemented in MATLAB as its toolbox provides built-in support for SIFT, SURF, HOG, GLOH in addition to the several classifiers. The hardware utilized comprised of a Core i3, 2.4 GHz processor with 4 GB Ram. The operating system used was Windows 8.0 64Bit and image processing techniques along with other related MATLAB toolboxes were employed in the development of the system. This research utilized 10-fold Cross-Validation for the analysis of

the data. It is a statistical technique used in evaluating predictive models by dividing the data into training and validation sets. The approach divided the dataset into K-fold equal data sizes and crossed over each K-dataset partition in successive rounds such that each data point has a chance of being in the validation process.

3.2.7 Result and Evaluation

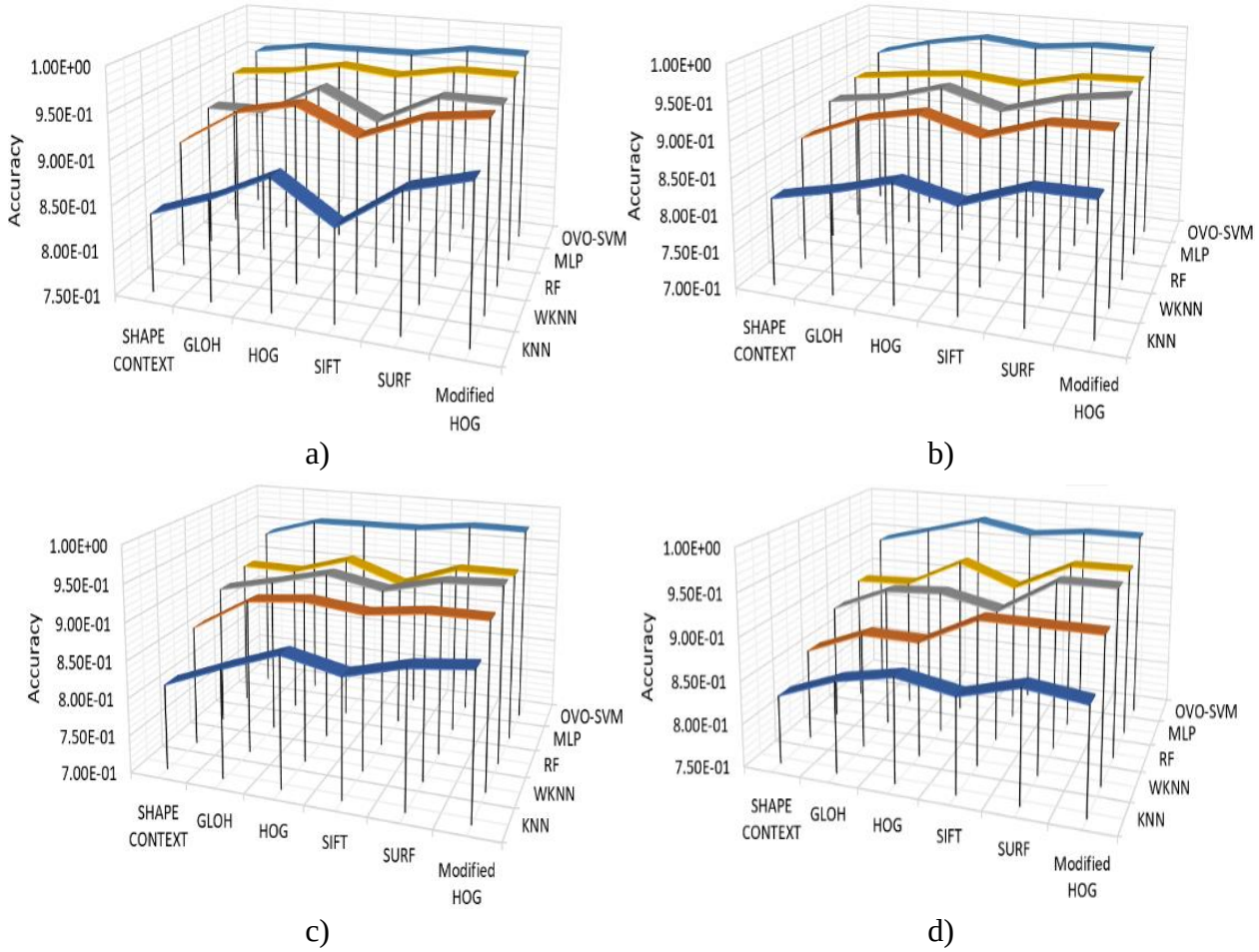


Figure 11. Accuracy measure of all dataset using feature descriptors on different classifiers.

Figure 11 measures the performance accuracy using SIFT, SURF, HOG, GLOH, Shape Context, and Modified HOG using KNN, WKNN, RF, MLP, OVO-SVM classifiers. The x-axis shows the accuracy while the y-axis gives the extracted features and the z-axis has the classifiers for the Naira, USD, CAD, and Euro Banknote. From the results, the best accuracy for the features and classifiers were SURF/Modified HOG and OVO-SVM, while the worst performance was achieved with SIFT and KNN respectively. Although the proposed Modified HOG presented

comparatively higher accuracy in most cases, it was barely higher than the SURF techniques.

Table 7. Error rate of feature reduction on descriptors using OVO-SVM on all dataset

No.	Descriptor	Feature Reduction	Naira	USD	CAD	Euro
1	**Shape Context**	No Feature reduction	0.0489	0.0627	0.0630	0.0594
		PCA	0.1500	0.1948	0.2104	0.1956
		BOW	0.1300	0.1553	1.7267	0.1678
		LVF	0.0950	0.1269	1.3056	0.1284
		HCF	0.0650	0.0754	0.0865	0.7253
		LVHCF	0.0700	0.0654	0.8721	0.6893
2	**GLOH**	No Feature reduction	0.0387	0.0421	0.0395	0.0415
		PCA	0.1300	0.1357	0.1243	0.1364
		BOW	0.0930	0.0892	0.7507	0.0885
		LVF	0.0600	0.0648	0.6340	0.0655
		HCF	0.0550	0.0540	0.5425	0.0574
		LVHCF	0.0575	0.0404	0.0409	0.0529
3	**HOG**	No Feature reduction	0.0379	0.0283	0.0379	0.0233
		PCA	0.1000	0.1092	0.1150	0.1144
		BOW	0.0900	0.0921	0.0988	0.1077
		LVF	0.0780	0.0795	0.0815	0.0921
		HCF	0.0550	0.0484	0.0502	0.0665
		LVHCF	0.0383	0.0285	0.0380	0.0238
4	**SIFT**	No Feature reduction	0.0371	0.0365	0.0368	0.0354
		PCA	0.1010	0.1204	0.1155	0.1288
		BOW	0.0750	0.0764	0.0843	0.0751
		LVF	0.0586	0.0582	0.0577	0.0549
		HCF	0.0400	0.0488	0.0422	0.0445
		LVHCF	0.0380	0.0469	0.0388	0.0363
5	**SURF**	No Feature reduction	0.0261	0.0276	0.0268	0.0275
		PCA	0.0907	0.1048	0.1166	0.1287
		BOW	0.709	0.0977	0.0928	0.0845
		LVF	0.558	0.0660	0.0702	0.0681
		HCF	0.350	0.0408	0.0502	0.0455
		LVHCF	0.0295	0.0389	0.3049	0.0315
6	**Modified HOG**	No Feature reduction	0.0266	0.0279	0.0268	0.0274
		PCA	0.0837	0.0884	0.0874	0.0865
		BOW	0.0649	0.0651	0.0662	0.0672
		LVF	0.0351	0.0257	0.0318	0.0328
		HCF	0.0266	0.0384	0.0296	0.0299
		LVHCF	0.0265	0.0277	0.0268	0.0276

Furthermore, RF and MLP did produce promising results. But their accuracy was dependent on the descriptor and currency as RF outperformed MLP in some cases and MLP outweighed RF in

performance accuracy in other feature and classifier combinations. Although, WKNN outperformed KNN, they both had low rates. Lastly, the lowest combination accuracy was SIFT and KNN with an average of 82%, and the highest accuracy combination was Modified HOG and OVO-SVM with an average rate of 97.3% with the proposed SURF closely following behind with an accuracy of 97.29%.

Additional experiments were conducted with feature reduction techniques to remove features that hinder the accuracy and precision of the proposed approach, and the error rate was utilized in measuring the performance of which is presented in Table 7. The table gives the Error Rate of OVO-SVM of the different Descriptors and Feature Reduction on the Four datasets. The proposed approach Modified HOG/LVHCF presented the lowest error rate within the average rate of 0.0270 after feature reduction. Meanwhile, shape context/PCA had the worse error rate with an average of 0.0585. The rankings of the descriptor and Feature reduction technique ranging from best to worse are Shape Context, SIFT, GLOH, HOG, SURF, Modified HOG and PCA, BOW, LVF, HCF, LVHCF respectively. Additionally, each feature reduction approach was applied to all the descriptors to determine the best pair with the highest recognition rate. From the result, PCA had the lowest acceptance rate across all datasets while Modified HOG/LVHCF presented the best results across all dataset with relatively higher performance than SURF. Regarding datasets, Naira note offered the best result while CAD had the worse rate although there were cases where CAD outperformed Euro and USD datasets when SIFT, SURF and Shape Context descriptors were employed. Lastly, the proposed approach presented promising results in comparison with existing techniques and can be engaged in descriptor dimension reduction.

Table 8. Experimental results of the counterfeit detection method

	Data Size	**Correctly Classified**	**Incorrectly Classified**	**Detection rate (%)**
Naira	450	418	6	92.89
USD	250	225	22	90
CAD	250	230	27	92
EUR	350	322	16	91.71

Table 8 presents the evaluation results of the proposed system on the datasets using OVO-SVM as it presented the best result during classification. The system was able to identify all images as banknotes even though it wrongly classified some bills. It showed that the Naira relatively had the highest correctly detected samples while the US dollar had the lowest successfully authenticated note. Although all datasets had few incorrectly detected samples irrespective of the size of the set.

The Naira had the highest detection rate, the Euro came second, the US Dollar was third, and the Canadian dollar had the lowest authentication rate.

For the multiple banknote recognition systems, an object detection method in a cluttered scene using point feature matching of extracted HOG features based on image patches computed from the vertices of SURF points normalized using low variance and high correlation filter. It reads the test image and preprocesses it by applying grayscale conversion, noise removal, background subtraction, and normalization. Also, the HOG Feature points are detected and used in matching the test image containing multiple banknotes is paired with the template image using the putative matching technique. The approach computes the transformation matched points while eliminating outliers. This aids in localizing the object in the scene. Additionally, the system locates the bounding polygon of the detected image and draws a black box to indicate the denomination.

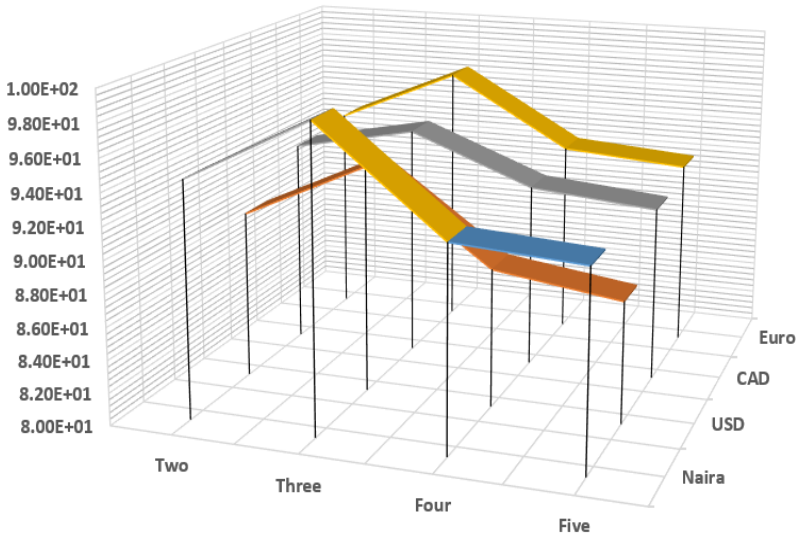


Figure 12. Recognition rates of a) Naira, b) USD, c) CAD, and d) Euro on multiple banknote

In Figure 12, we have measured the Recognition Rate using one, two, three, four and five notes on the Naira, USD, CAD, and Euro Database. For simplicity, we have shown here only the recognition rate on the x-axis, while the number of the banknote is fixed on the y-axis and the currency type on the z-axis. The reason for this experiment is to evaluate the effectiveness of the proposed feature extraction approach on multiple banknote recognition. The system recognized several currencies including same note in the same image, separated, overlapped, cluttered and partially obstructed. After three bills the acceptance result reduced when the number of banknotes was increased. In all cases, the accuracy measure is more than 87%. Although one, two and three

notes show more than 92% while four and five ranged between 87.50% and 92.45%. For this research, the best recognition rate across all dataset was three bills using Naira at 98.60%, and the worse was with five banknotes using USD at 87.50%.

Table 9. Qualitative analysis of feature descriptors

Method	Time	Scale & Rotation	Illumination	Affine
Shape Context	Better	Better	Common	Good
SIFT	Common	Best	Common	Good
SURF	Best	Common	Common	Good
HOG	Common	Common	Best	Best
GLOH	Good	Best	Common	Good
Modified HOG	Best	Better	Best	Best

In Table 9, we empirically compare the performance of all feature descriptors for multiple banknote identifications concerning its variants in five distinct scenarios: speed, scale, rotation, illumination, and affine invariance. The results of the experiment were analyzed qualitatively to give a general idea of the performance of each algorithm in distinct scenarios using four grades: Common, Good, Better, and Best. As shown in the table above, Shape context has significant advancements in speed, scale, affine, and rotation changes; yet not in illumination change. Furthermore, SIFT is slow; but is invariant to rotation and scale, while maintaining a stable degree of stability for affine transformation and a low threshold to illumination change. In the same view, SURF is fast and has better performance than SIFT and GLOH, but it is not stable to the rotation and illumination changes. HOG is robust to rotation, illumination, positioning, and affine modifications but very slow and scale variant. Although GLOH has excellent performance in scale and rotation, presented reasonable speed and affine changes and had a deficient sense of illumination invariance. However, the Modified HOG descriptor maintains a proper stability illumination, rotation, and affine transformation. It can also handle reasonable scale invariance and compared with the other algorithms; it has distinct advantages regarding speed. Thus, can be deduced that the proposed technique presented a promising performance in banknote recognition. Lastly, the result of this qualitative analysis is not constant in all cases as changes to the dataset, and classifiers can lead to a different conclusion.

One critical issue is determining the number of notes by denomination while authenticity is needed to ensure a reasonable performance from the proposed technique. To solve this problem, the various databases are evaluated concerning the sizes by varying the training set dimension. A

remarkable increment on the recognition and error rate is seen when the training dataset increases from 300 to 800 for banknote recognition, 100 to 250 for counterfeit detection, and 50 to 200 for multiple currency recognition. Growing the data further only produced asymptotical performance improvement, but had no effect on its accuracy and error rate.

Chapter 4

4 Mobile Based Banknote Reader

4.1 Mobile App for the Detection of Counterfeit Banknote

The mobile based banknote recognition system was designed to recognize the face value of banknotes captured using a smartphone camera. The App was entirely based on a mobile device and did not need any additional hardware as it was dependent on the processing capability of the smartphone. Finally, the system can be used on either side of the banknote and outputs detection results in text and audio. Fig. 13 presents the flow diagram of the banknote reader. The reader acquires the banknote image, activates the banknote recognition system and then outputs the value or/and authenticity of the bill.

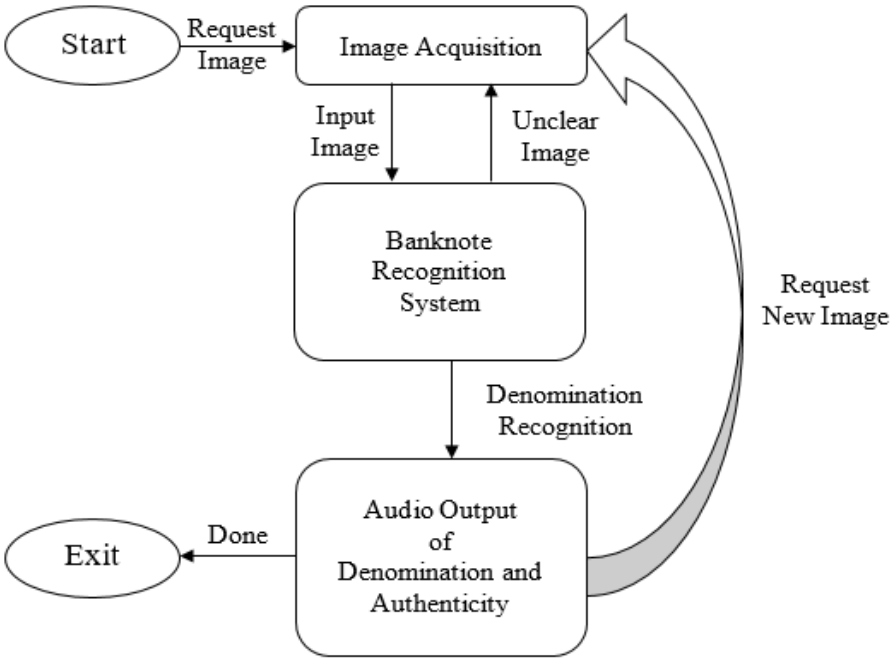


Figure 13. Flow diagram of banknote reader

4.1.1 Image Acquisition

When the Mobile app is launched, the system provides two options for banknote acquisition. In the first choice, the user selects an existing Naira image on the smartphone. With the second one, the user activates the cell phone camera and takes a snapshot of the note for recognition. This phase includes an object tracker module to handle banknote detection issues that occur when using this system.

4.1.2 Region of Interest (ROI) Extraction

The first step carried out during this phase was separating the Region of Interest (ROI) from the rest of the image. The banknote is converted into grayscale, and ROI extraction is achieved by dividing the image into partitions. The partition containing the ROI is extracted using the detected edges. Computation time is reduced by using the knowledge that the face value can only be found in two specific partitions in a banknote. We optimized the system by configuring it to check just two partitions, depending on the orientation of the Naira note. It then locates the face value by identifying all the blobs in these partitions and selects the most prominent blobs of the banknotes by cropping and utilizing it as the face value.

Also, the extracted ROI contained noise that was reduced using a combination of gamma correction and histogram equalization to enhance the image contrast. Furthermore, the filtered ROI is then binarized, and stored in a matrix with *W* representing a white (background image) pixel and *B* for a black (banknote value) pixel [105]. Additionally, Slant correction using the image histogram is employed. The approach computes the horizontal and vertical projection histograms, shears it by the angle that maximizes the height of peaks in the vertical projection of the figures, uses the Zhang-Suen Thinning algorithm. Finally, the image was normalized to uniform the dataset. The result of the region of interest extraction is shown in Fig. 14 which presents a distribution of good to worn out banknote contained in the dataset.

Figure 14. Extracted ROI image

4.1.3 Feature Extraction

The feature extraction process involves removing features from the preprocessed image to examine and categorize the distinctive and peculiar characteristics of each banknote denomination to classify our note image. In this research, 16 features were extracted and combined as feature vectors: projection, crossing, distance, direction, geometric moment, variance, skewness, kurtosis, eccentricity, orientation, Euler number, centroid, area, intersection, start, and endpoint.

4.1.4 Classification

In the classification system, the first phase involves constructing the base classifiers using the five distance measures of the KNN method to obtain a very good generalization of the banknote predictions. Four distance measurements were utilized namely: Euclidean, Manhattan, Bray-Curtis, and Canberra distance where *x* and *y* are vectors in n-dimensional space in equations 10,11,12, and 13 respectively [23]. During the process, the lapses of the existing KNN variants were identified. Therefore, a variation to adapt to the errors led to the development of a new Distance Measure.

The proposed similarity measure method computes the absolute variance between the extracted features of the template image and the test image using equation 14. The differences are multiplied together, and the various products are compared. The class with the lowest value is selected since it has the closest distance to the new image. The multiplication of the absolute variance was introduced to avoid cases where two or more classes have an equal number of images nearest to the test image. Also, the KNN phase includes a benchmark for rejecting images that are not recognized as a Naira notes.

$$\text{Distance (x, y)} = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2} \quad (10)$$

$$\text{Distance (x, y)} = \sum_{i=1}^{d}|x_i - y_i| \quad (11)$$

$$\text{Distance (x, y)} = \frac{\sum_{i=1}^{d}(x_i - y_i)}{\sum_{i=1}^{d}(x_i + y_i)} \quad (12)$$

$$\text{Distance (x, y)} = \sum_{i=1}^{d}\frac{|x_i - y_i|}{|x_i + y_i|} \quad (13)$$

$$\text{Similarity (x, y)} = \sqrt{\prod_{i=1}^{d}|x_i - y_i|} \quad (14)$$

For further evaluation, the results of the best three distance measures were then combined using the Stacking and Weighted Majority Algorithm (WMA) ensemble learning methods that led to the proposing of a Cascaded Ensemble algorithm. Weighted Majority Algorithm (WMA) is a heterogeneous ensemble learning algorithm employed in constructing a compound algorithm from a pool of prediction algorithms. Furthermore, the technique is a binary decision problem and builds a compound algorithm by assigning a positive weight to each base KNN classifier. It then computes the weighted votes of all the base models in the pool and finally assigns the sample to the prediction with the highest majority. Lastly, the WMA technique assumes no prior knowledge of the detection rate of the KNN algorithms in the pool but instead believes that one or more of the base classifiers will perform well. While stacking involves an ensemble classifier learning to combine multiple KNN based classifiers generated by different distance measures on a single dataset, the approach produces a set of base-level classifiers to train a linear regression classifier to combine the outputs of the base KNN models.

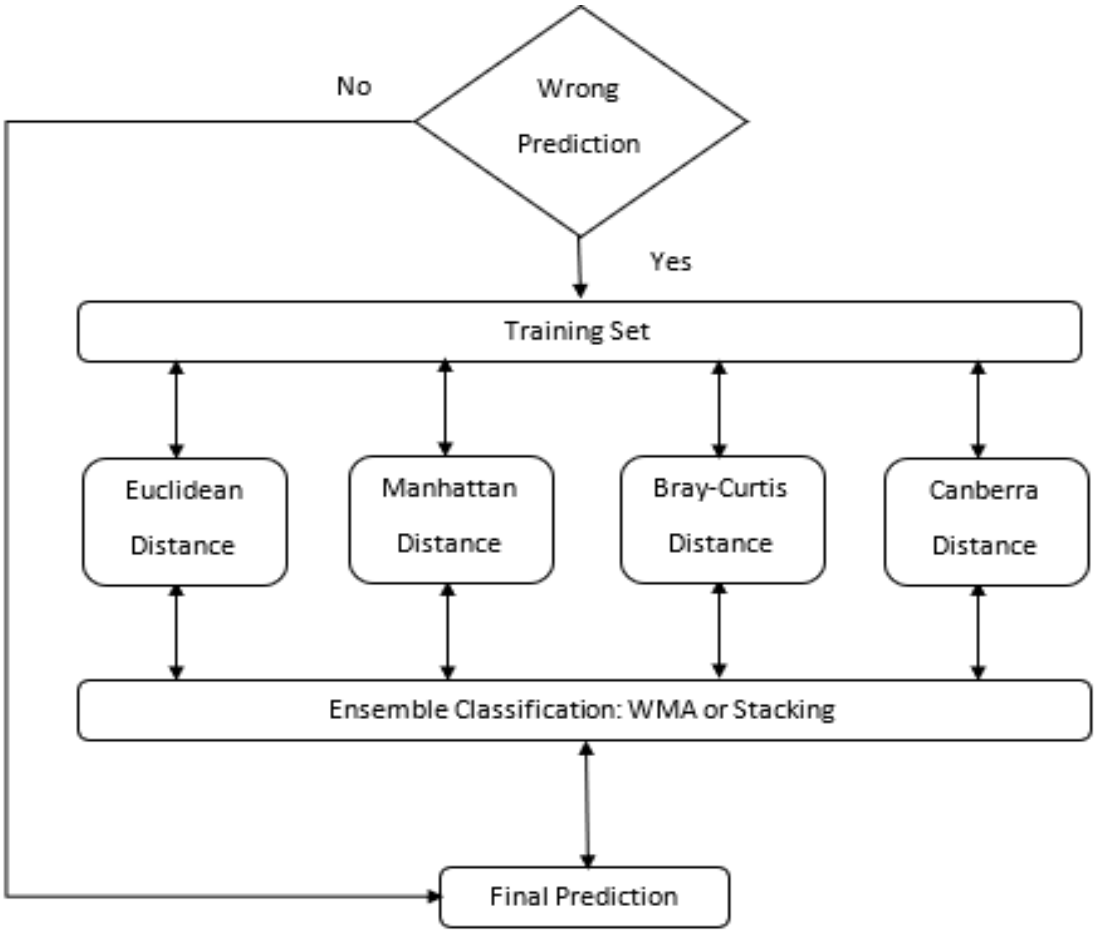


Figure 15. Proposed ensemble KNN algorithm

Furthermore, the individual classifiers were trained using the whole training dataset; before linear regression is applied to fit using the outputs of each distance measure in the ensemble [42]. The Cascaded Ensemble approach employs a two-level classification; the first phase is based on the distance measure of the KNN classifier with the highest accuracy (true positive) and recall (false negative). Which is used in the classification/detection of the banknote by relating the unknown data to the known data based on a calculated distance or similarity measure. Meanwhile, the second stage employs a WMA and Stacking algorithm as meta-classifiers by using the remaining KNN methods as the base classifier. The flowchart of this approach is presented in Figure 15.

4.1.5 Data Collection

The dataset images were captured by using two acquisition methods. One uses a Canon T2i digital camera while the second uses an iPhone 6. Pictures were taken from different angles. The digital camera images were used to recreate physical banknote taken with the mobile phone. The iPhone 6 captured real-time photos used for the training and testing processes. The Naira note was gathered from existing Nigerian banknote from the north, south and eastern parts of the country. They were collected and labeled by the CENPARMI research group and contained different distortions like stains, mutilations, fading, and tears representing the reality of circulated paper money in Nigeria. Ten-fold cross-validation was employed for analyzing the dataset. For this research, 2310 Genuine and 2048 fake Nigerian banknotes were collected ranging from Five Naira to One Thousand Naira (₦5, ₦10, ₦20, ₦50, ₦100, ₦200, ₦500 and ₦1000) and the system had 16 classes: 8 for the genuine note and 8 for fake bills.

4.1.6 Experimental Setup

This research utilized Unity 3D as its development platform. It is an application designed by the Danish Unity Technologies Company in 2005. It consists of a fully developed multimedia system with built-in features like rendering, illumination, behavior, audio, video and sound provisions. Unity 3D also has a powerful user interface development facility. It is supported by multiple platforms (Android, iOS, and Windows) and deployed on mobile phones, computers, virtual reality systems, gaming consoles, web applications, and smart TVs. It also bypasses the use of conventional programming languages like C++ and Java, in addition to working around

platform dependent applications like Android or iOS and their related toolboxes. Two platforms were chosen for this work: Android using Samsung S6, and iOS via iPhone 6S. Lastly, developing a solution in Unity 3D is quick, since the Unity Editor has many powerful convenience tools and has an integrated Asset Store with many free or easy-to-purchase plugins for any system. The proposed approach evaluates the viability of completely designing a mobile-based banknote recognition and authentication technique rather than scaling down existing systems to be compatible with the processing capability of mobile devices. The KNN and Ensemble classifiers were constructed on the Unity platform from scratch rather than scaled down from other existing software. During the process, the lapses of the current KNN and Ensemble variants were identified. Therefore, a variation to adapt to the errors led to the development of a new Distance measure and a Cascaded Ensemble based approach. The Unity system is multiplatform, so we tested the system on the Android and iOS platforms using a Samsung Galaxy S6 and an iPhone 6 while recording the average recognition time. The application was designed as a dedicated banknote recognition system running on a smartphone that communicates the recognition output to the visually impaired by using a pre-recorded text and verbal message.

4.1.7 Result and Evaluation

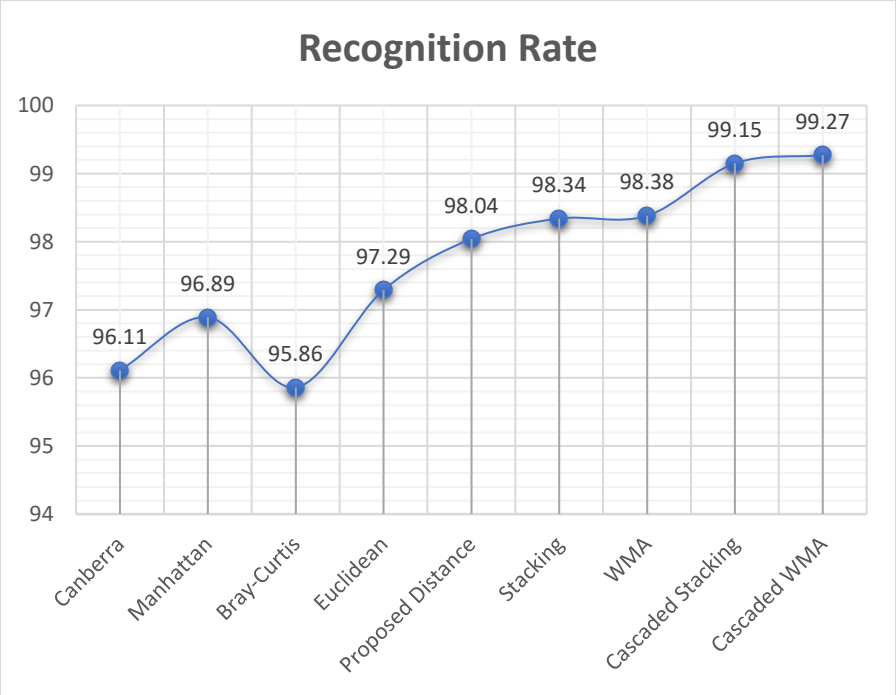


Figure 16. Recognition rate (%) of classifiers

Figure 16 presents the recognition rates for Canberra, Manhattan, Bray-Curtis, Euclidean, Proposed Distance Measure, Stacking, WMA, Cascaded Stacking, and Cascaded WMA. In general, all distance measure achieved relatively high recognition rate. However, Bray-Curtis had the lowest recognition rate of 95.86%, and our proposed Cascaded WMA Ensemble Classifier presented the highest recognition rate of 99.27%.

Furthermore, Table 10 displays the existing state of the art mobile-based systems trained using our dataset. Although, some existing research also achieved comparative results without outperforming the proposed method. They had considerably longer processing time as they either used a client-server architecture or were tailored down to fit the computational capability of mobile devices. Finally, these previous works are yet to extend their approach to incorporate authentication of genuine and fake notes.

Table 10. Comparison of the proposed method with state of the art

Distance Measure	**Technique**	**Recognition Rate (%)**	**Recognition Time**
Liu 2008	AdaBoost Framework	97.30	0.11 Second
Singh et al. 2014	Bag of Words Sift Descriptor K-mean	95.95	1.08 Second
Doush et al. 2016	SIFT KNN	72.4	72.6 Second
Proposed Method	Face Value (OCR) Cascaded Ensemble Classifier	99.27	0.02 millisecond

In the same view, Fig. 17 shows the detection rate of the eight Nigerian banknote denominations using Canberra, Manhattan, Bray-Curtis, Euclidean, Proposed Distance Measure, Stacking, WMA, Cascaded Stacking, and Cascaded WMA. The lowest denomination detection was recorded in the 1000 Naira note using Bray-Curtis distance while the highest was also achieved on 1000 Naira note by our Cascaded Ensemble Classifier. Furthermore, all distance measure performed better with note 100 Naira and above as they contained three individual characters while 50 Naira and below included one or two distinct characters had a lower recognition rate. Additionally, the proposed distance measure outperformed all other existing techniques employed in this research.

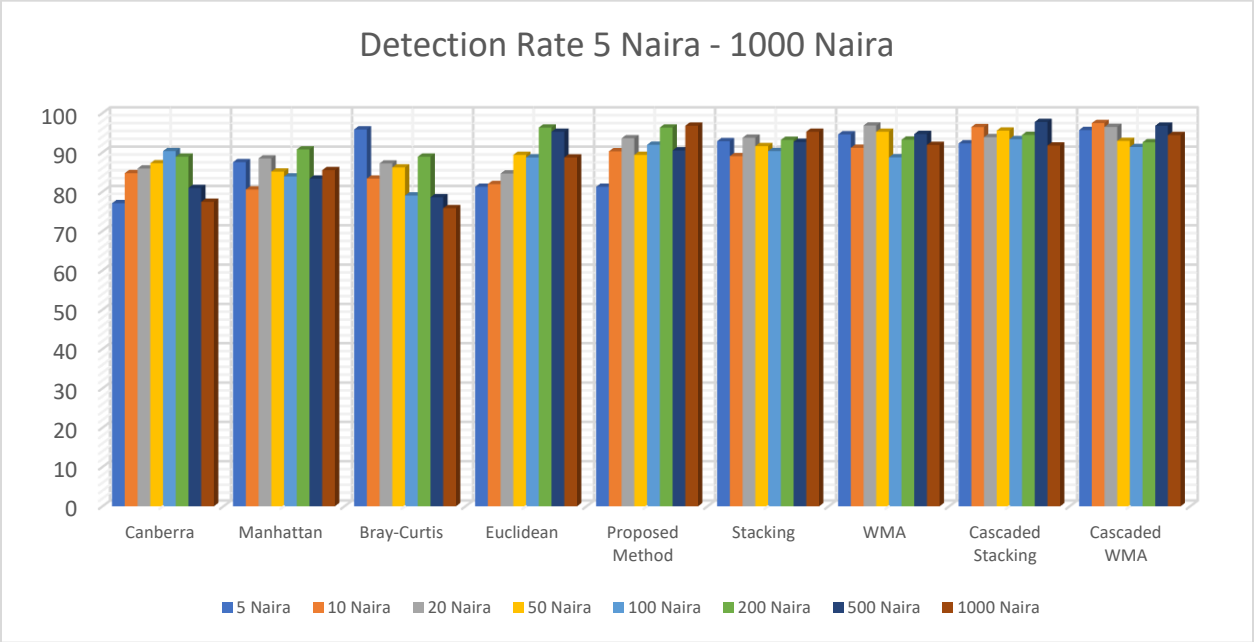


Figure 17. Detection rate of banknote on class/distance measures

Figure 18 presents the detection rate for Canberra, Manhattan, Bray-Curtis, Euclidean, Proposed Distance Measure, Stacking, WMA, Cascaded Stacking, and Cascaded WMA (Classifier) utilized in this research. Canberra had the lowest detection rate of 84.06% while our proposed distance measure offered the highest detection rate of 91.24%. Also, in Figure 19 images of genuine and fake banknotes that were rejected by the system are shown. Thus, this study presents a mobile application for paper currency identification and authentication using OCR feature in conjuncture with an adaptation of the Nearest Neighbor and Ensemble Classifier. The proposed banknote recognition system had a 99.27% recognition rate, a detection rate of 94.70% and processing time of 0.02 millisecond.

During the classification phase, the mobile app first checks to see if the face value of the test image meets the set threshold value, which was based on the average face value feature set. If the feature set used does not fit the image threshold, it gives an audio and text output requesting the user to recapture or select the image again, in addition to providing the necessary steps needed to ensure the system works correctly. The system also has a help menu that gives a detailed explanation and guidance in audio and text format informing users of the steps to go through for the recognizing and authenticating of banknotes.

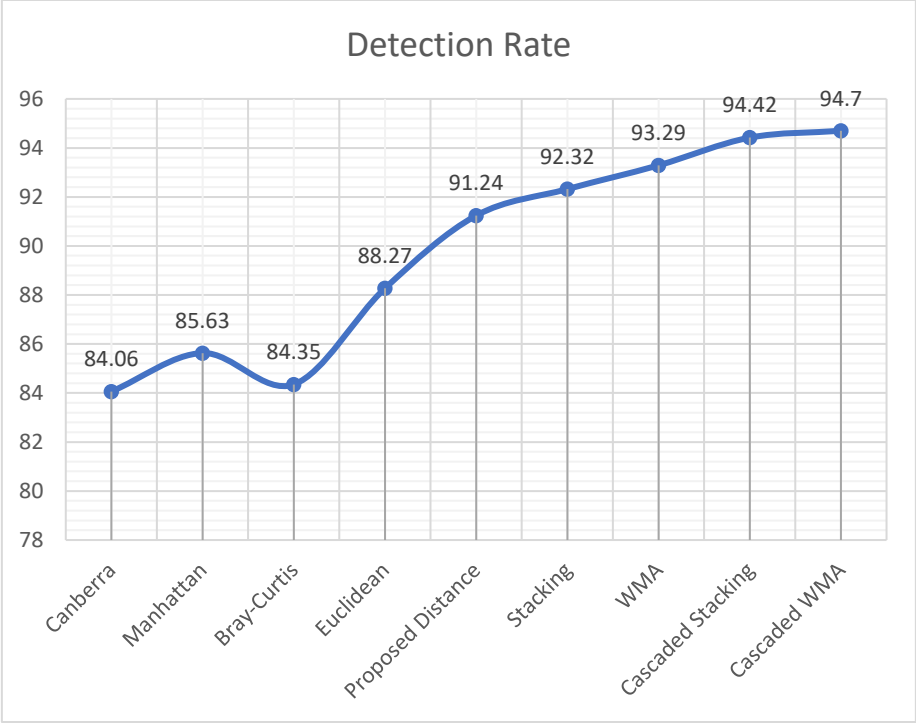


Figure 18. Detection rate (%) of classifiers

To utilize the system for another currency, the banknote image can be uploaded into the system via existing pictures or by activating the mobile phone camera on the application. If bulk training/testing needs to be done, then the existing image option is utilized. However, if a single image training or testing approach is required, the mobile phone camera is activated to capture the note in real time. Moreover, the system must be configured to select the partitions that include the face value of the currency of choice. The system can also be configured to use all or some of the extracted features. The text input can be updated for any currency of choice, and since the system employs a text to audio converter upgrading the recognition output text to banknote of choice automatically updates its audio output to reflect the new currency being tested by the mobile application.

a)

b)

Figure 19. Rejected a) genuine, and b) fake banknote

4.2 Mobile Based Assistive Technologies for Authentication of Banknotes

The primary purpose of this research is to identify original or forged banknotes and their values. A framework for authentication of paper currency based on the analysis and experimentation using SIFT and SURF keypoint based features classified using a stacked MLP and ECOCSVM is implemented. Figure 20 depicts the SIFT/SURF based Money Reader system. The system conducts image acquisition and preprocessing on the mobile phone. Then, the image is transmitted to a MATLAB backend for feature extraction and classification, later the results are sent back to the cell phone for text and audio output of the denomination and its authenticity evaluation.

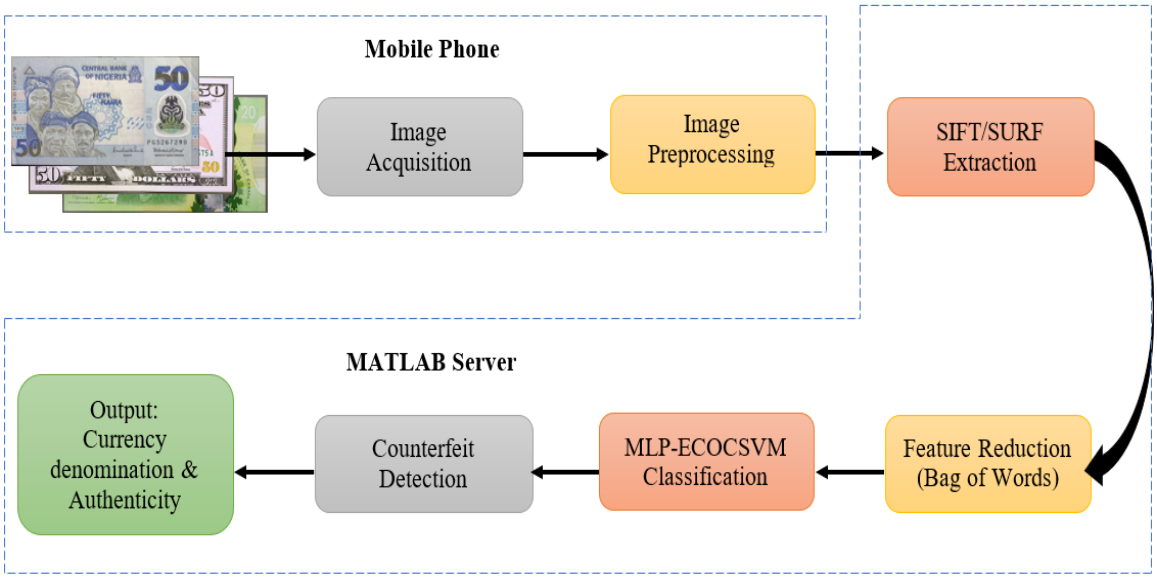


Figure 20. The proposed SIFT/SURF money reader

4.2.1 Image Acquisition

The database consists Nigerian notes, US dollar and Canadian dollar bills with different years of issue. The images were acquired using a digital camera and were stored in JPEG format. The photos are in a 24-bit color image format with a resolution of 300 dpi (Dots Per Inch), and each database had two faces: obverse and reverse.

4.2.2 Image Preprocessing

After the banknote images have been digitized, the next step is preprocessing. It involves converting to grayscale then segmentation of the banknote by separating the bill from its

background. The second level requires improving the contrast of the image by applying gamma correction and histogram equalization. By using edge detection, the boundary coordinates of the banknote are cut out of the original image [20]. The banknote is then rescaled to ensure uniformity in note sizes. Figure 21 shows a sample pre-processed image transmitted from the mobile phone to a MATLAB backend server; a) consists of full note without occlusions, rotations and illumination changes, b) partial note with lighting changes, distortions and clutters, and c) distorted banknote with wrinkles and occlusions. Furthermore, the first row consisted of Naira note, the second row comprised of US Dollar while the third row included Canadian Dollar.

a) b) c)

Figure 21. a) Full note, b) partial Note, and c) distorted note

4.2.3 Feature Extraction

Feature extraction is a critical phase in the banknote classification and detection process. To efficiently classify and authenticate a note, it is crucial to define its features and characteristics adequately [36]. This research employs a combination of two keypoint based methods known as Speeded-Scale Invariant Feature Transform (SIFT) and Speeded up robust features (SURF) technique.

SIFT Feature Extraction: This method retrieves descriptors by fast-tracking the localization of the keypoints while preserving the property of the image [17]. It is a feature extraction method for detecting and extracting local and distinct features. The technique identifies keypoints in the image, including the blobs, edges, and corners in the banknote. The system locates the keypoints in the regions of the Naira note using equation 15. *Kmli* is the number of keypoints which is achieved by the summation of the keypoint identified in the jth region (j = 1, 2 ... m) of the banknote using the 1st descriptor in the ith number of classes of the naira. Each SIFT detector located several keypoints from different regions of the Naira note [80].

$$Kmli = \sum_{j=1}^{m} Kjl \quad 15$$

For each key-point in the image, a feature descriptor is computed from the magnitude of the gradient and its relative direction of the pixels in the local neighborhood [19]. The method groups the pixels into 16 x 16 sub-regions and subregion further bundled into 4 x 4 regions in the keypoint. Each sub-region is made up of an 8-dimensional weighted directional histogram, eight bins, each turned by 45° [11]. The number of feature vectors (*FVi*) in the note of ith classes can be obtained using equation 16 [34].

$$FV_i = \sum_{l=1}^{p} K_{mli} \quad 16$$

SURF Feature Extraction Description: It is based on a local invariant fast feature point detector in addition to a distinctive feature point descriptor, which is rotation and scale invariant. It employs the Hessian-based blob detection method in locating interest points and is created from the sum of Haar wavelet responses [80]. SURF computes the second order Gaussian filter for all keypoints (x, y) in an image, then applies a Hessian matrix Happrox at a scale σ, where the Happrox is given by equation 17.

$$Happrox = \begin{bmatrix} L_{xx}(a,\sigma) & L_{xy}(a,\sigma) \\ L_{xy}(a,\sigma) & L_{yy}(a,\sigma) \end{bmatrix} \quad 17$$

To facilitate the filtering based on the second order Gaussian filter, a convolution approach is employed by using the different scales based on a square-like filter where det (Happrox) denote the highest value in the image pyramid [12]. In order to assign the best orientation to each keypoint, a sliding orientation window is then rotated around the keypoints. The Haar wavelet transform is used to compute the weight of each pixel in the given window using equation 18.

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad 18$$

To aggregate the vectors, an area made up of 20x20 pixels is created and further grouped into 4x4 area called block, thus using a 3-dimensional vector to generate the feature descriptor [136]. Figure 22 presents the preprocessed image with detected SIFT and SURF keypoint descriptors of Naira, USD and CAD paper money employed in this research.

Combined SIFT/SURF Descriptor: This research also implemented a feature combination technique based on the SIFT and SURF features. The approach detected SIFT and SURF features from the banknote separately, computed the descriptors at each identified keypoint, normalized

the SIFT and SURF descriptors and then concatenated them creating a COMBINED SIFT/SURF descriptor employed as feature vectors used in the recognition and detection of the bills.

Figure 22. Detected SIFT and SURF descriptors of a) Naira, b) US, and c) Canadian dollar

4.2.4 Feature Reduction Using Bag of Words

After the feature vector had been constructed, it was realized that the dimension was significant and would increase the processing time. The inclusion of unstable descriptors could also affect the recognition rate. Thus, K-Mean was introduced to create a cluster of related features. It is an unsupervised learning process that analyzes the distribution of samples in the feature space [47]. The technique groups random samples from the SIFT, SIFT and COMBINED feature vector descriptors. The value of *K* is set to the number of related feature groups. The K-mean algorithm is then run until an average result is reached. Finally, the BOW model is utilized to represent the clustered features by computing the visual words to reduce the number of feature vectors [83]. For experimental evaluation, additional analysis was conducted using Information Gain (IG) and PCA in reducing the feature set.

4.2.5 Classification

Classification involves determining the similarities and dissimilarities of a representative dataset to find quite smaller data dimensionality of the represented data [92]. For this research, the system employed stacking in Combining a multilayer perceptron (MLP) neural network with Error-Correcting Output Codes Support Vector Machine (ECOCSVM). Stacking is an approach used in training a learning algorithm to unite the predictions of different classifiers. MLP is an interconnected collection of nodes where each node denotes artificial node and arrow signifying a connection from the output of one node to the input of the next layer. The neural network comprises the following layers: input, hidden and output layers [44]. On the other hand, the ECOCSVM technique compares all classes against each other and builds a binary classifier to distinguish each

pair of class while ignoring the remaining classes. When classifying an unknown image to a particular class, the classifier conducts a vote, and the class with the majority is chosen as the best choice to assign the banknote [41]. First, the dataset is split into training and testing sets, MLP and ECOC are employed in training and testing the dataset. Furthermore, a combination mechanism based on a single layer linear regression model is utilized to make a final prediction using the results of the MLP and ECOCSVM as inputs and the correct responses as the outputs.

4.2.6 Data Collection

The dataset consisted of 3110 Naira note, 1943 US Dollar, and 1733 Canadian Dollar which were further divided into genuine and counterfeit bills. The dataset of each currency is divided into original and fake paper currencies. The Naira note has eight classes, which are ₦5, ₦10, ₦20, ₦50, ₦100, ₦200, ₦500, and ₦1000. While the US dollar has six categories, namely $1, $5, $10, $20, $50, and $100, the Canadian Dollar has five groups, namely $5, $10, $20, $50, and $100.

4.2.7 Experimental Setup

In this paper, we bypass the use of conventional programming languages like C++ and Java or platform dependent application such as Android or IOS and their relating toolbox. We instead focus on developing the mobile phone front end of the system using a Multiplatform framework called Unity 3D. It utilizes JavaScript and C# as its scripting language and can build executable applications that can run on the Web, Windows, Android, Flash, and IOS. Due to its scalability, ease of development and powerful processing capability, it is employed in developing mobile applications [129]. While on the backend, MATLAB 9.2 R2017a was employed in the implementation of the modules. The decision to use MATLAB was because of the extensive mathematical functionality of this package as it enables inexperienced users to work with the toolbox in addition to providing automated and batch standardization of analyses and statistical tools for data representation [77]. The system was tested on an iPhone 6 and a Samsung S6 with a minimum camera resolution of 8 megapixels and had a processing speed of 1.4 GHz.

This research presents a banknote recognition method based on the SIFT/SURF combination approach that classifies and authenticates banknote using a stacked MLP and ECOCSVM. For further evaluation, another classification approach was applied to determine the robustness of the system. The SIFT and SURF keypoint descriptors have been utilized on computers, but due to its

computational complexity, it cannot be fully migrated to a smartphone. Thus, we present a hybrid method that preprocesses the image on the smartphone before sending it to a MATLAB server for SIFT and SURF feature extraction and classification using a stacked MLP and ECOCSVM.

4.2.8 Result and Evaluation

Table 11. Feature reduction set

	Feature Reduction Set			
	Feature Set	**IG**	**PCA**	**BOW**
SIFT	513434	480	540	500
SURF	342918	480	540	500
COMBINED	856352	620	780	650

Table 11 shows the feature reduction techniques and results employed in this paper. The SIFT and SURF method presented an average feature set of 513,434 and 342,918 vectors. After feature selection, IG had 480 features, PCA gave 540 vectors, and BOW introduced 500 features respectively. With the COMBINED approach, the concatenated descriptors had 856,352 vectors, IG produced 620 features, PCA had 780 vectors, and lastly, BOW presented 650 features. It was evident that we could achieve a feature reduction ratio near 90% along with a decrease in classification time (about 55%) comparing with the non-optimized features.

Figure 23 is a comparative analysis of the feature reduction approach on four classification methods on the Naira note. It was introduced to reduce the processing time in addition to diminishing the feature set and equalized the feature vector amongst all classes and database. The system was classified without feature reduction implemented, and for further evaluation, IG, PCA, and BOW were employed to reduce the SIFT, SURF, and COMBINED descriptors. However, there was not any significant difference in accuracy between the feature vectors without feature reduction and vectors after BOW feature reduction. Moreover, it can be deduced that the proposed MLP-ECOCSVM method outperformed Naïve Bayes Multinomial (NBM), MLP, and ECOCSVM. The feature reduction and classification approach that presented the best accuracy was SIFT/SURF combination (BOW) and MLP-ECOCSVM while SIFT/SURF (IG) and KNN had the lowest accuracy. Lastly, SURF technique outperformed SIFT with or without feature reduction using IG, PCA, and BOW.

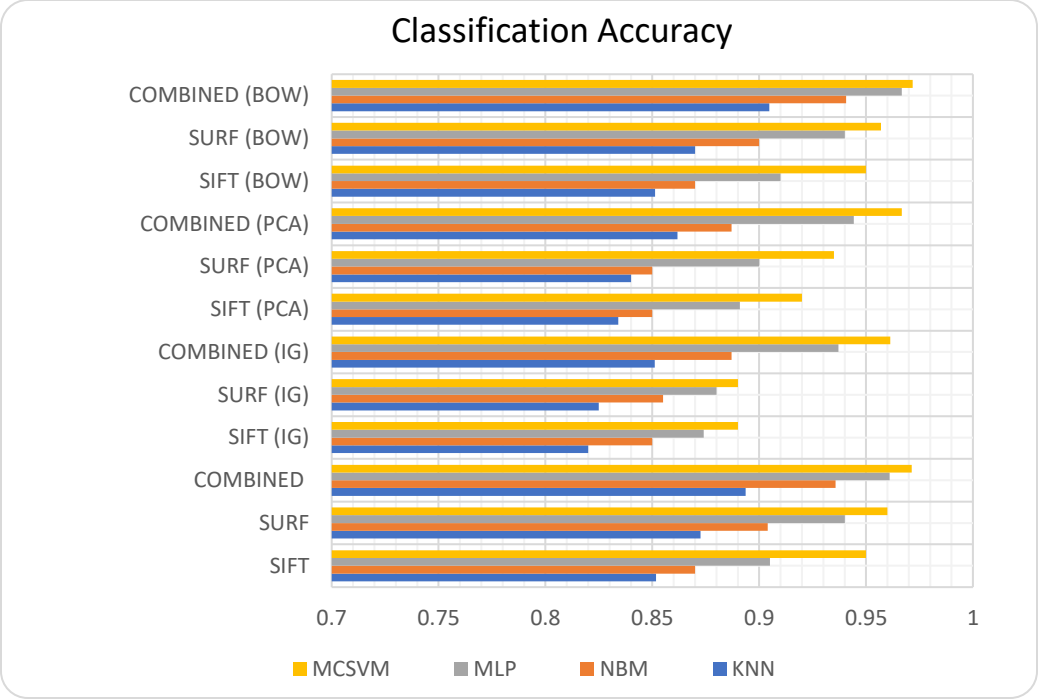


Figure 23. The accuracy of feature reduction and classification technique on Naira dataset

Figure 24 displays the result of the performance criteria utilized in this research. The Accuracy, Precision, Recall, and F-Measure of the proposed COMBINED-BOW and MLP-ECOCSVM me thods on Naira, USD, and CAD Database are shown. The result proved that the keypoint based feature (SIFT and SURF Concatenation) method employed in this research could be useful in practical banknote recognition systems. However, the Naira note database presented the highest Precision, Recall, Accuracy, and F-Measure while Canadian Dollar bills had the lowest results.

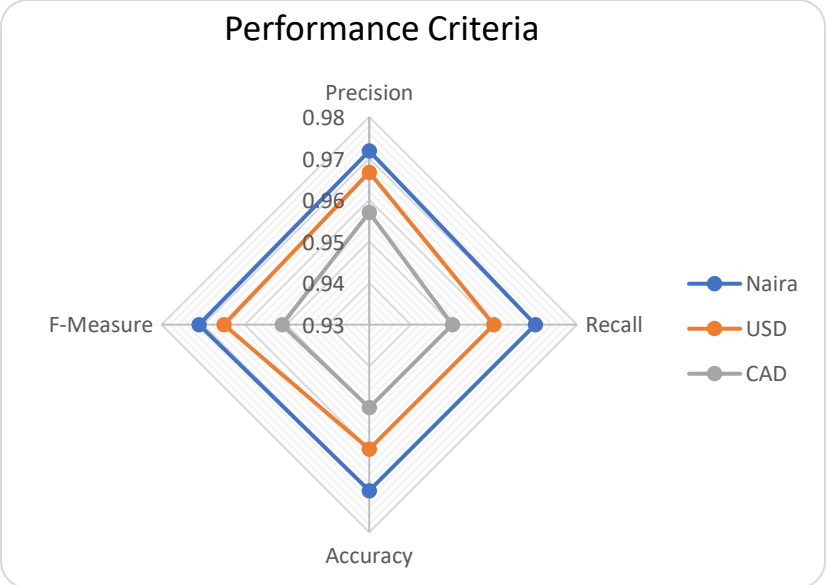


Figure 24. Performance criteria of the proposed technique on MCSVM on all dataset

In Figure 25, the recognition time is computed for SIFT, SURF and COMBINED using IG, PCA and BOW feature reduction methods using the proposed approach on the Naira, USD, and CAD databases. From the result, the SIFT and SURF outperformed the combination across all feature reduction methods; although the difference was minimal. Additionally, PCA and IG could reduce the recognition time; but BOW had the best recognition time. However, the COMBINED had the longest recognition time due to the size of the feature vectors. Lastly, the Canadian dollar database produced the best processing time while Naira had the worst processing time when Bag of words reduction is employed.

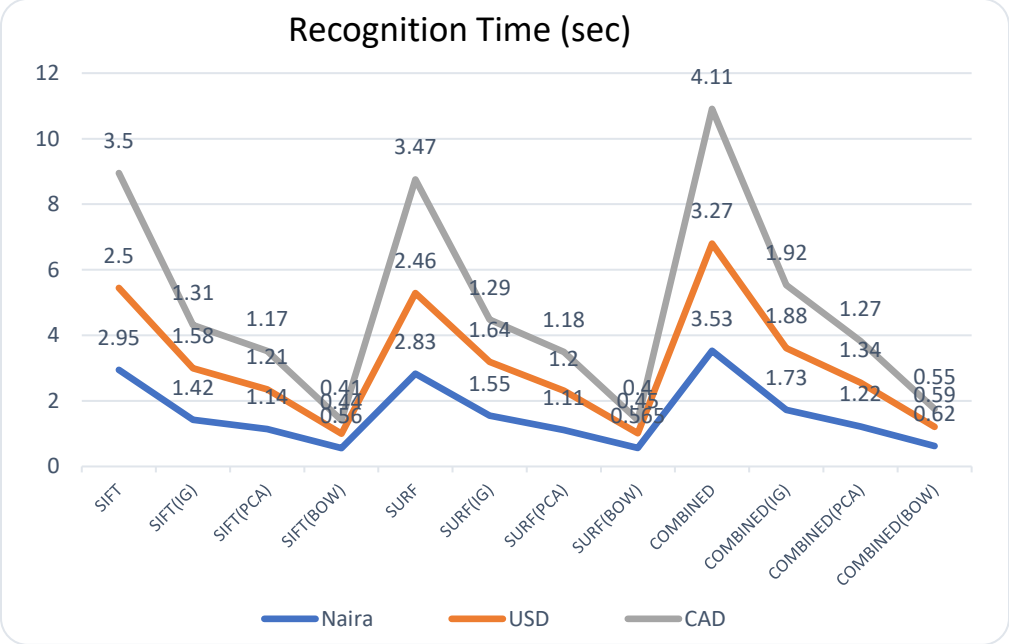


Figure 25. Recognition time (sec) of Naira, USD, and CAD

Table 12 portrays the Detection Rate (DR) and Error Rate (ER) of SIFT, SURF and the COMBINED approach using several feature reduction methods on MLP-ECOCSVM classification technique. Although COMBINED (BOW) gave the best result, SURF(BOW) outperformed the other Feature reduction approaches and presented competitive detection and error rates. Finally, SIFT(IG) had the lowest detection and error rates across all dataset utilized in this research.

Moreover, the experimental evaluation shows that our proposed approach presented a competitive Recognition rate and lowest Error Rate irrespective of the quality of the fake banknote dataset. Several other feature extractor/descriptor combinations have also been tested in this research: FAST, ORB, and FREAK. However, they produced a lower recognition/detection rate,

extended the processing time or presented insignificant improvement in the system accuracy. To reduce the processing time of the system an Apache server using a PHP 5 and a MATLAB automated service is set up. It allows an instance of the MATLAB script to be always in memory thus eliminating boot time delays. Furthermore, the method communicates with the mobile phone using a COM protocol through PHP, making it easy to access the server. The smartphone, on the other hand, utilizes a POST request approach in sending banknote images to the server while the server streams back binary data to the mobile phone after recognition/detection of the banknote.

Table 12. Detection and error rates of classifiers with the MCSVM

	Naira Note		**US Dollar**		**Canadian Dollar**	
Classifier (%)	**DR**	**ER**	**DR**	**ER**	**DR**	**ER**
SIFT	93.00	7.00	92.56	7.44	92.26	7.74
SIFT(IG)	85.44	14.56	84.85	15.15	85.89	14.11
SIFT(PCA)	90.00	10.00	87.71	12.29	90.57	9.43
SIFT(BOW)	93.77	6.23	93.00	7.00	93.11	6.89
SURF	94.60	5.4	94.46	5.54	93.45	6.55
SURF(IG)	87.89	11.11	86.96	13.04	86.34	13.66
SURF(PCA)	92.00	7.00	90.82	9.18	91.49	8.51
SURF(BOW)	94.45	5.55	93.75	6.25	92.32	7.68
COMBINED	95.11	4.89	94.58	6.42	94.99	5.11
COMBINED(IG)	92.34	7.66	93.45	6.55	93.67	6.33
COMBINED(PCA)	94.23	5.77	94.02	5.98	94.25	5.75
COMBINED (BOW)	95.45	4.55	94.75	5.25	95.32	4.68

4.3 High Correlation-Based Banknote Gradient Assessment of Ensemble Classifier

The primary purpose of this research is the identification of banknote denomination and counterfeit detection. The system is based on the analysis and experimentation with the pixel features extracted from grayscale banknote images by the Histogram of Gradient (HOG) method. Four feature reduction approaches were applied to reduce the feature set that is then classified using Weighted Majority Average (WMA) ensemble classifier with Multilayer Perceptron Neural Network (MLPNN) and Multi-Class Support Vector Machine (MCSVM) as base classifiers. Due to its computational complexity in feature extraction, it cannot be fully migrated to a smartphone. Thus, we present a hybrid method that preprocesses the image on the smartphone before sending it to a server for feature extraction, dimensionality reduction, and classification. The front end was

designed using Unity 3D while the backend is programmed using MATLAB, and it includes a user-friendly database comprising Naira, USD, CAD, and Euro banknote. Figure 26 depicts the phase of the proposed paper money recognition system.

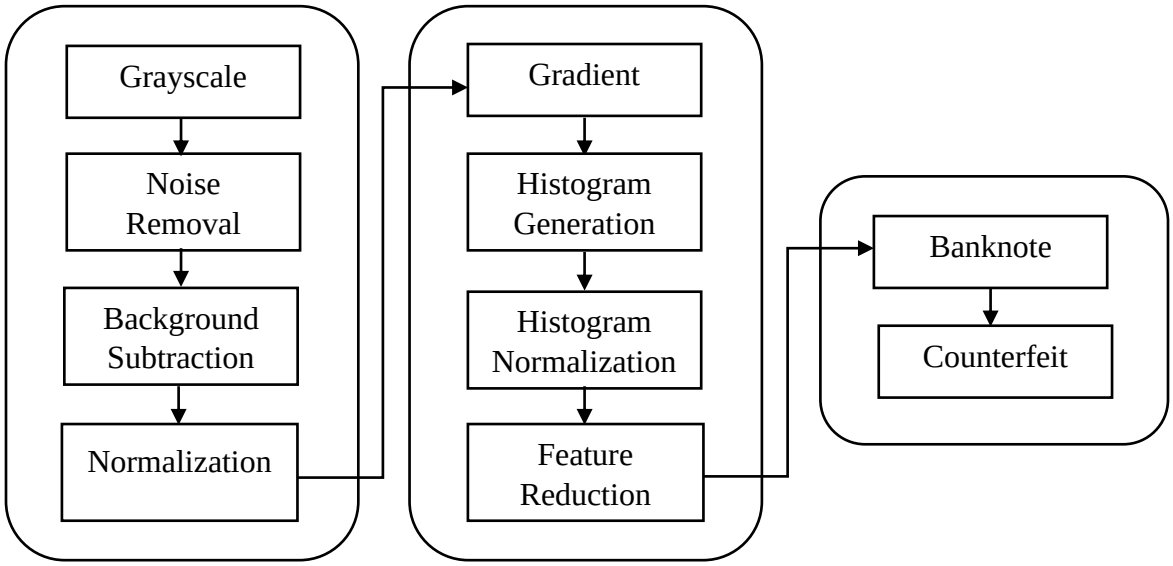


Figure 26. Block diagram of the proposed banknote recognition system

4.3.1 Image Preprocessing

Image pre-processing is a fundamental step in recognition systems; it involved enhancing and cleaning up the banknote images for efficient feature extraction. The pre-processing methods utilized were grayscale conversion, noise removal, background subtraction, and normalization; they were all implemented on the front end using Unity 3D. Figure 27 shows a sample pre-processed image transmitted from the mobile phone to a MATLAB backend server drawn from the four datasets used, namely Canadian dollar (first Row), Euro note (second Row), US Dollar (third Row), and Naira (last Row). a) comprised of full banknote without clutter, rotations, obstructions, and lighting changes, b) partial banknote with distortions, rotations, and illumination changes, and c) occluded banknote with wrinkles, obstructions, and distortions.

Grayscale Conversion

To convert the images to grayscale, we first calculate the average of the three channels of the RGB image. The result is represented as a matrix whose fragments are assigned values based on how dark or bright the color of the pixel is at the corresponding position.

Noise Removal

Digitized banknote images sometimes have discontinuous particles and noise. In this research, a Gamma Correction and Weiner filter were applied to reduce the amount of noise present in the image.

Background Subtraction

During image acquisition, the irrelevant background image is captured. In order to recognize the banknote, this unwanted data must be removed. Background subtraction is achieved by blob detection. The technique pinpointed the boundaries of the note in the digital image where the brightness changed to separate it from its background image and was able to distinguish weak blobs.

Normalization

Banknote denominations vary in sizes, and the note can also change due to the acquisition manner. When matching the image, if the banknote is of varying dimension, it could affect the recognition rate. Thus, we rescale all note to the same height and width.

a) b) c)

Figure 27. a) Full note, b) partial note, and c) occluded note

4.3.2 Feature Extraction

To efficiently recognize, classify and retrieve a banknote, it is very imperative to define its features and characteristics adequately. In this research, four different descriptors were employed: shape context (SC), Scale Invariant Feature Transform (SIFT), gradient location and orientation histogram (GLOH), and Histogram of Gradient (HOG).

Shape Context

Shape Context (SC) is a 3-dimensional histogram of edge point locations and orientations that computes the distance and orientation histogram of other points relative to the interest points using

a Canny edge detector [13].

Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) constructs a scale space by reproducing an integral image that is smaller than the original image [72]. Furthermore, a detection method is employed in computing the distance between the scales. Interest points are then located by analyzing the blobs, edges, and corner by finding the highest and lowest pixels in each neighborhood. After which, a matrix is utilized to compute the principal curvatures, and a detection method is being used to remove interest points below the threshold [123].

Gradient Location and Orientation Histogram

Gradient Location and Orientation Histogram (GLOH) were developed as an extension to the SIFT descriptor to improve its robustness and distinctiveness. This technique extracts a set of interest points. It constructs a scale space then applies a Gaussian method to compute the distance between the scales and localizes the interest points. Additionally, an orientation is assigned to each interest point to get rotation invariance. Finally, Histogram of Gradient descriptor is computed, quantized into bins and PCA is employed to reduce the dimension of the data [77].

Histogram of Gradient

The Histogram of Gradient (HOG) descriptor partitions a banknote image into small evenly-dimensioned sections called cells comprising several pixels which individually have weights computed by getting the gradient magnitude and histogram [132]. The histogram shows the dispersion of the pixel taking into cognizance the gradients orientation. Lastly, cell histogram is offered as a vector that is merged to get the HOG feature of the overall banknote image [68].

4.3.3 Feature Reduction

To reduce the processing time several feature reduction methods have been implemented: Principal Component Analysis (PCA), Bag of Words (BOW), and High Correlation filter (HCF).

Principal component analysis

Principal component analysis (PCA) is a statistical approach used to reduce the extracted feature dimension while retaining trends and patterns. It employs a mathematical process that converts some possibly correlated features into a smaller number of uncorrelated features called

principal components. It clusters highly correlated features while making sure it retains the highest possible variance and unusually diminishes the least square reconstruction error [122]. The process starts by normalizing the size of the dataset, computing the covariance matrix of the dataset that is employed to construe the Eigenvectors and Eigenvalues. The Eigenvalue is arranged in descending order, and k eigenvectors corresponding to the most substantial eigenvalues are chosen where k is the scale of the new feature subspace ($k \leq dk \leq pd$). The projection matrix w is created from the chosen k eigenvectors. Lastly, the original data x is transformed into w using the k-dimensional feature subspace Y [18].

Bag of Words

Bag of Words (BOW) is a feature reduction approach adapted from the field of natural languages processing. It employs the matrix formed after extracting feature vector created from each image in a dataset. Bag of words builds a visual vocabulary by reducing the feature dimension by clustering the feature subspace using K-means clustering. This approach provides an encoding technique for counting the visual word occurrences in an image and produces a histogram that becomes the new and reduced representation of the banknote feature vector [102].

Low variance filter

Low variance filter (LVF) is a dimensionality reduction approach for measuring the amount of variance in the feature vector extracted from each image in a database. The technique computes the variance of each feature vector and deletes vectors with variance value below a set threshold. The low variance filter can only be used with feature vectors with numerical values as the value of the variance is dependent on the numerical range of the feature set. Therefore, the vectors need to be normalized before calculating their variance to make the values independent of the dataset domain range. For this research, all feature vectors are normalized ranging from 0 to 1. The vector variance is calculated before the vectors below the threshold 0.04 are removed. All remaining feature sets are then de-normalized to return to their original numerical value. Finally, any threshold higher than the stipulated 0.043 employed in this research gave a lower accuracy value proving that feature reduction is not only necessary for execution time but also performance improvement.

High Correlation Filter

High Correlation Filter (HCF) is based on the relationship between the extracted feature set. It computes the intra-feature values variability to ascertain the similarity between the features to

remove points with high correlation as they do not add any new information to the existing feature pool. Additionally, High Correlation Filter (HCF) was achieved by using a linear correlation node to estimate the similarity between correlated feature pairs by employing the Pearson's Product Moment Coefficient. To measure the relationship between the nominal values, a threshold of 0.269 is applied to remove any one of the two highly correlated feature points.

4.3.4 Classification

The classification stage is the decision segment of a recognition system, and it uses the features extracted in the previous phase. It involves identifying each banknote and assigning it to the correct detection class. Furthermore, the principal function of a pattern recognition system is to give decisions concerning the class membership of the models with which it is confronted. This paper employs a Weighted Majority Average (WMA) ensemble classifier to combine Multilayer Perceptron (MLP) Neural Network and MCSVM as base classifiers to create a strong classifier. It is a heterogeneous ensemble learning method used in constructing a compound process from a pool of prediction algorithms. The WMA technique assumes no prior knowledge of the detection rate of any of the base techniques but instead believes that one or more of the base classifiers will perform well. The MLPNN can be trained to estimate almost any smooth, quantifiable function with no assumption on the division of data. The technique can easily model linear and nonlinear functions and can be trained to generalize if presented with new and unseen data. While the MCSVM system uses a binary classification approach for the one versus one technique, it designs a classifier for each class and creates a hyperplane between the classes while assigning a class to a test sample based on a distance margin for the classes.

4.3.5 Data Collection

The banknote recognition and counterfeit detection dataset designed by the CENPARMI research group contained Naira Note, US Dollars, Canadian Dollars, and Euro Bills. The real banknote was acquired using a Canon T21 camera at a close distance in a controlled environment and normalized to a 200dpi color; while the fake note was then generated using a 1200 dpi HP Color LaserJet CP4520 printer. Each dataset had two faces: obverse and reverse and was split into a two-fold set to realize a two-fold experiment each employed to one of the base classifiers.

4.3.6 Experimental Setup

This research proposes a hybrid mobile phone application built using Unity 3D for front-end development. Unity 3D is a useful development tool with the ability to work on both the iPhone and Android-based platform while possessing the capacity to improve mobile phone functionality and resolved the unpredictable growth hassles of mobile phone applications. Besides, it gives an application developer the opportunity to scale many of the hurdles faced as it supports several scripting languages like C# and JavaScript and is not platform independent. Developing a solution in Unity 3D is quick as it has many powerful convenience tools and integrated asset store with many free or easy-to-purchase plugins for any system. While the backend employed MATLAB 2018a because of the extensive mathematical functionality of the package. It enabled inexperienced users to work with the toolbox in addition to providing automated, and batch standardization of analyses and statistical tools for data representation [95]. Lastly, the system accepts BMP, JPEG, TIFF, and PNG image formats.

4.3.7 Result and Evaluation

Table 13 displays the precision and recall of Naira, Euro, CAD, and the US note employed in the banknote recognition experiments using SC, SIFT, GLOH, and HOG with WMA (MLPNN+MCSVM). This was experimented to evaluate the feature extraction method using our proposed Ensemble Classifier in recognition of banknote and to determine which of the several techniques can best predict the note denomination. From the result, the HOG technique presented the best feature vector results for this experiment while SC had the lowest recognition rate across all database. USD banknote had the worse result across all methods and Naira gave the best precision and accuracy.

Table 13. The precision and recall results of classification using WMA (MLPNN + MCSVM)

Method	**Naira**		**USD**		**CAD**		**Euro**	
	Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
GLOH	95.40	95.40	92.30	91.96	93.46	93.00	94.28	94.28
SIFT	94.56	94.55	90.44	90.24	92.95	92.74	93.34	92.31
HOG	97.60	97.40	94.10	94.1	95.53	95.52	96.72	95.96
SC	92.45	92.38	88.56	88.56	91.56	91.56	92.29	92.28

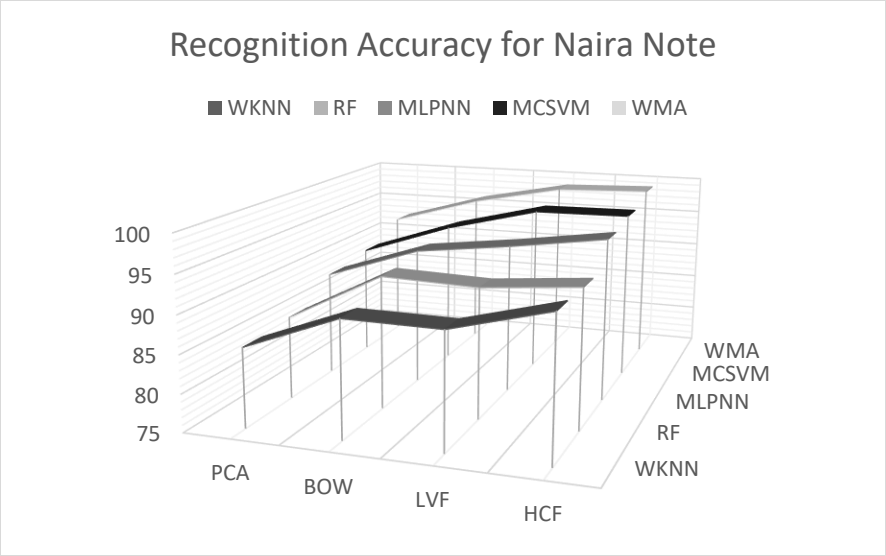


Figure 28. Naira note recognition accuracy

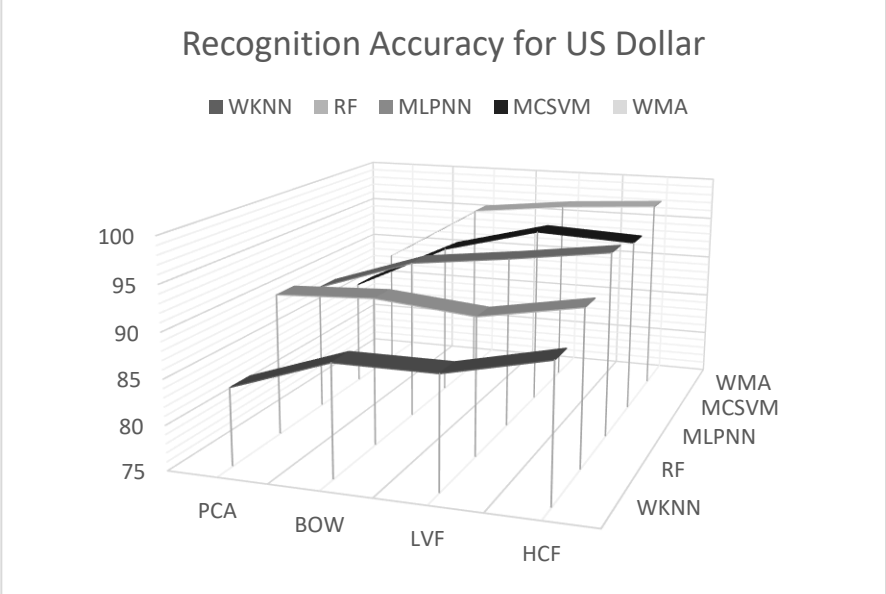


Figure 29. US dollar recognition accuracy

Figure 28, 29, 30, and 31 present the recognition Accuracy of Naira, Euro, USD, and CAD banknote respectively. The Feature reduction method employed is PCA, BOW, LVF, and HCF while the Classifiers used are WKNN, RF, MLPNN, MCSVM, and WMA. Furthermore, the best feature reduction and classification approach employed was HCF and WMA, and the worse methods were PCA and WKNN. The ranking of the feature reduction method and classifiers ranging from best to worse are HCF, LVF, PCA, BOW and MCSVM, MLP, RF, WKNN, and WMA. Each feature reduction approach was applied then to all the classifiers to determine the best

pair with the highest recognition rate. From the result, PCA/WKNN had the lowest acceptance rate across all datasets while HCF/WMA presented the best recognition across all dataset although its result is not significantly higher than LVF/MCSVM. Regarding dataset, Naira note offered the best outcome while CAD had the worse rate despite there were cases where CAD outperformed Euro and USD dataset when MCSVM and MLP Classification approach was employed. From the recognition rate, HCF and WMA were adopted as the principal feature reduction and classification approach used in this work.

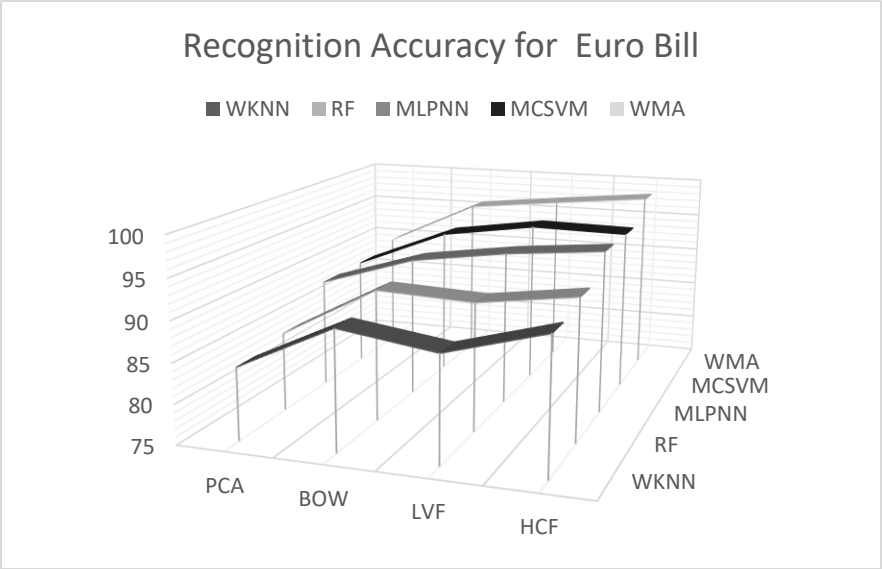


Figure 30. Euro bill recognition accuracy

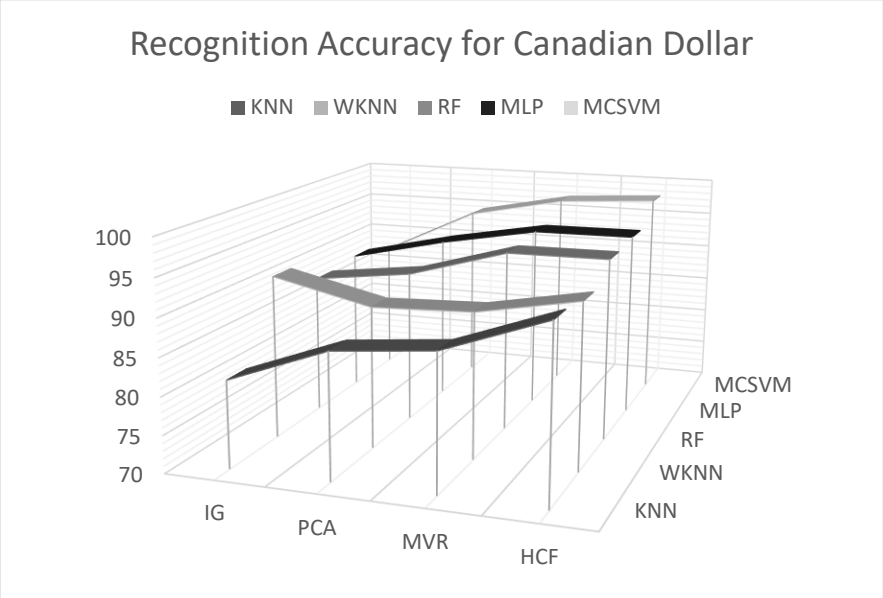


Figure 31. Canadian dollar recognition accuracy

Figure 32 presents the processing time for each dataset employed in using HOG/HCF technique on the WMA ensemble classifier method. The pre-processing image time for the datasets differed broadly as the maximum period was with the Naira dataset and minimum US Dollar dataset respectively. The pre-processing of each image varied because of the difference in image size, quality, and pixel mass. During the feature extraction, the Canadian bill had the shortest processing time, while the US dollar had the longest. Each dataset extracted a different number of features hence the disparity in timing. In the same view, the training time for all dataset was relatively close. The testing time for all dataset was in the same range although the Canadian dollar had the lengthiest time. From the processing time, we deduced that the banknote dimension caused a variation in the image processing time and the feature extraction time also dispersed as a different number of interest point detectors, descriptors; pixels, cells, and blocks utilized in each dataset. However, the training and testing, the differences were small irrespective of the dataset.

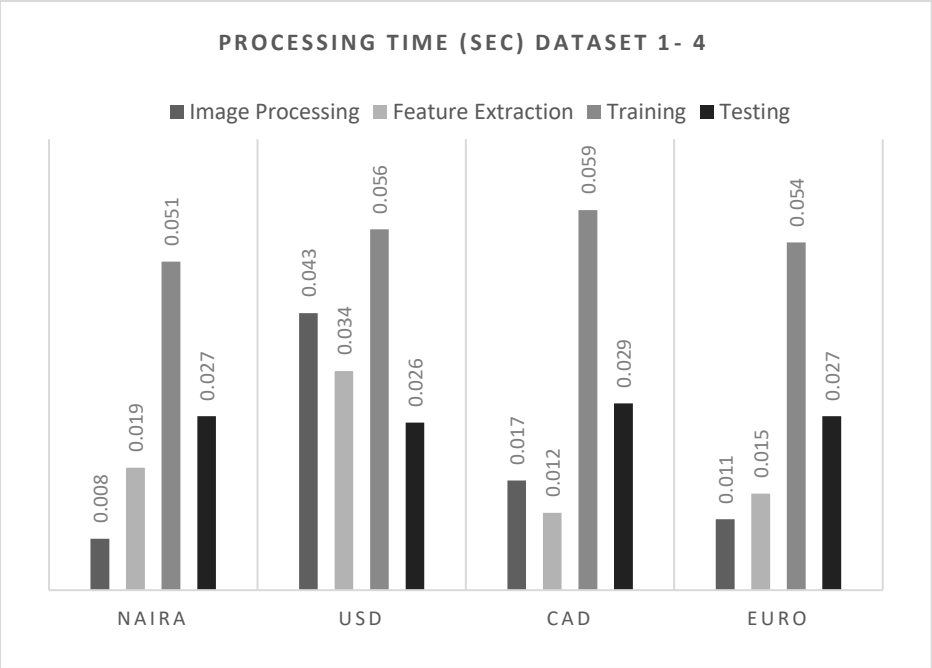


Figure 32. Efficiency regarding processing time (sec) of (HOG-HCF and WMA) proposed system

Figure 33 shows the detection rate of GLOH, SIFT, SC, and HOG after applying the HCF feature reduction technique on the proposed WMA Ensemble Classifier using different dataset. Although the detection rate of the individual feature vectors was very promising, the proposed approach when HOG descriptor is extracted achieved the best detection rate. Furthermore, SIFT and GLOH had a slightly lower rate to our proposed HOG while SC had poor results with a low

detection rate. For the individual dataset, USD had the lowest performance while Naira had the best with CAD and Euro slightly behind with a barely significant difference in accuracy.

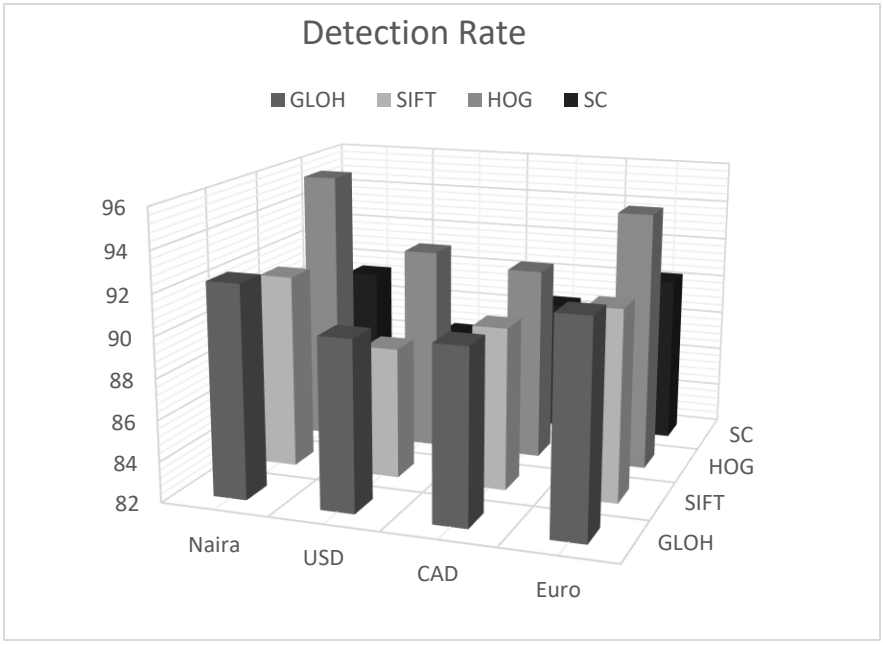


Figure 33. Detection rate of proposed system

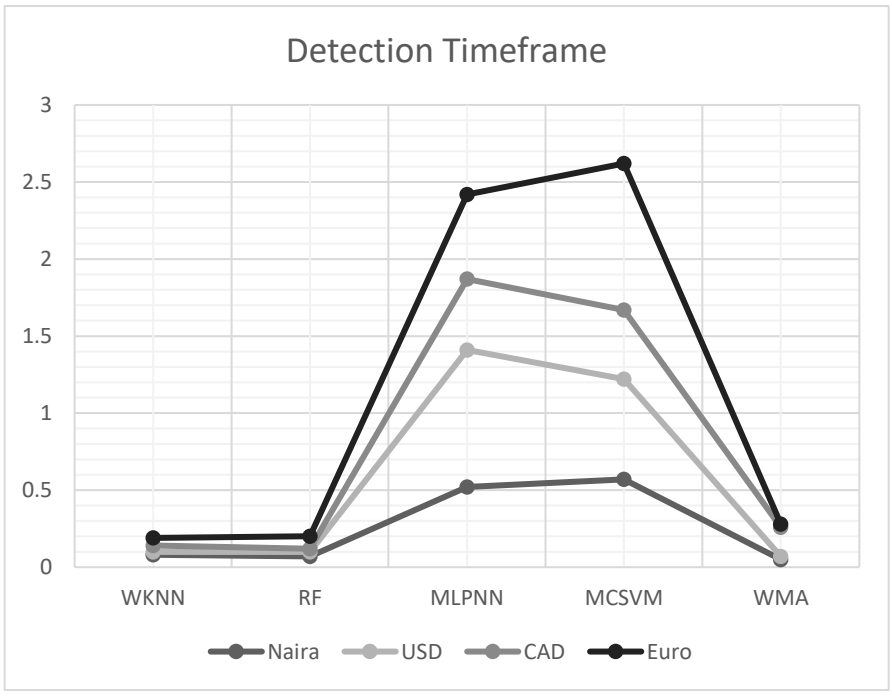


Figure 34. Average detection time

Figure 34 shows the average detection time for each dataset with each classifier. The Euro and USD had the shortest processing time when classified using the WKNN, and WMA, while the CAD presented the shortest detection time with the RF, MLPNN, and MCSVM. The Euro and USD note had the shortest time on the proposed classification method and CAD had the highest processing time. From the result, we concluded that the banknote dimensions caused a variation in the image processing time and the feature extraction time was dispersed as a function of the number of pixels, cells, and blocks utilized in each dataset. The feature reduction process was taken into consideration as it contributed to the processing time. Although, the validation time differences for each classifier were small regardless of the datasets. In conclusion, although all other classifiers have a remarkable detection rate and computation time, the suggested WMA classification method produced more promising results although each image varied because of the difference in image size, quality, and pixel density.

Table 14. Comparison of the proposed method with state-of-the-art techniques

	Technique	**Recognition Rate (%)**	**Recognition Time (Seconds)**
Liu 2008 [69]	AdaBoost Framework	97.50	0.10
Paisios et al. 2011 [86]	SIFT K-mean	88.45	0.15
Singh et al. 2014 [108]	BOW SIFT K-mean	96.70	1.09
Doush et al. 2016 [32]	SIFT KNN	74.52	72.40
Proposed Method	HOG-HCF WMA	99.00	0.051

Furthermore, Table 14 shows a comparison of the performance of the proposed system with the existing state of the art mobile-based systems. The techniques employed with the state-of-the-art approach include SIFT, BOW, KNN, K-mean, and Adaboost. They were all tested on the Naira, USD, Euro and CAD dataset used in this research. The suggested algorithm has improved the detection accuracy by more than 2% on average compared to previous feature-based detection algorithm. Although some existing research achieved comparative results without outperforming the proposed method, they all presented considerably longer processing time. Finally, these previous works are yet to extend their approach to incorporate authentication of genuine and fake note.

To minimize the recognition time of the Mobile-based banknote reader a PHP 5 and MATLAB automated service was set up using an Apache server, this ensured an instance of the MATLAB script to always be in memory, mitigating boot time delays. Additionally, the system transmitted banknote image to the server using a POST request while a COM protocol is utilized to communicate back to the mobile phone with both protocols streaming binary data. The administrator of the system can also edit the input text which automatically updates the audio as the App utilizes a text to audio converter that gives an audio description of the currency being tested by the mobile application. Lastly, the system includes a help menu that provides detailed guidelines in audio and text informing users of the phases involved in recognizing and authenticating the banknote.

Firstly, most mobile phone cameras have issues focusing on the necessary close distance needed to capture the security features of a banknote or when the banknote is placed far away from the camera. The resolution of the acquired image diminishes and could hurt the recognition rate of the system. Secondly, the camera is small and tend to produce more noise, thus affecting the quality of the captured image. Thirdly, smartphones have limited computational capability, hence the need to keep the computational cost within the scope of the device. Finally, apart from the challenges of the device limitation, the image acquisition process also poses some further issues that can have a negative impact on our banknote recognition systems such as the banknote orientation, position, background and lighting conditions.

Chapter 5

5 Mobile Based Ensemble Classification of Deep Learned Feature for Medical Image Analysis

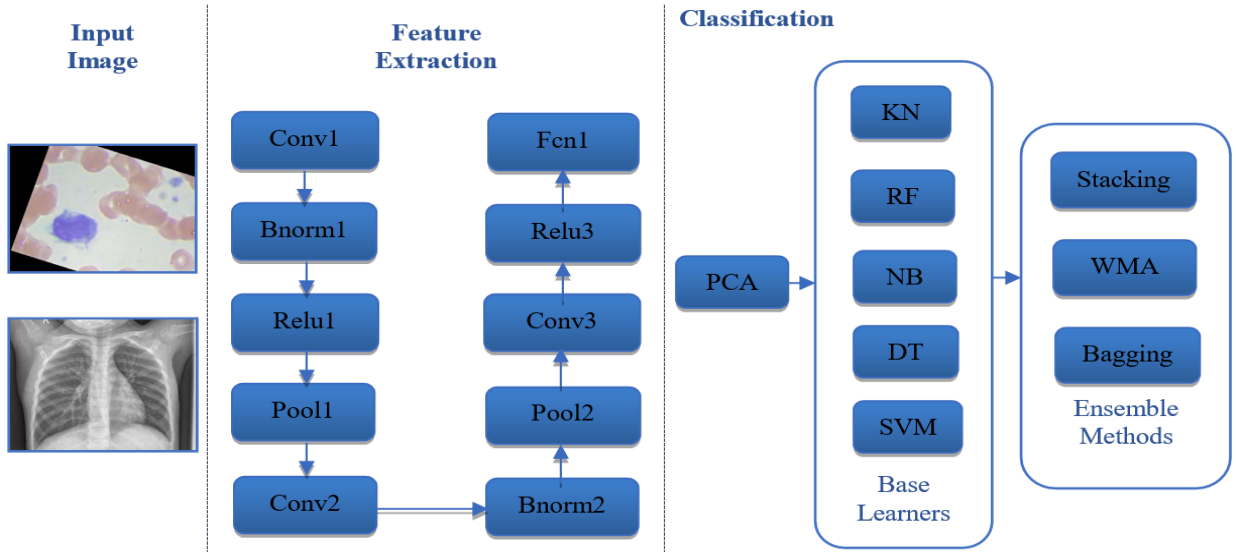


Figure 35. The architecture of the convolutional neural network ensemble

This research proposes a pre-trained CNN ensembled model that exploits the outstanding descriptive capability of deep neural networks for extracting features. It also takes into consideration the appropriate dimensionality reduction, classification and ensemble learners that can best utilize the structure of the deep learned feature space. In this section, we first describe the dataset, hardware, and software employed in the study, followed by the proposed CNN Ensemble as the definition of the input data and desired outputs before the actual methods to provide a better definition of the problem and a better understanding of the technique. The method pre-classifies the medical images, so the training is conducted on each class to learn the different network models needed to classify blood cell and chest x-ray samples. The medical image analyzer acquires the medical image using the Unity 3D frontend and, performs image formatting, feature learning, quantization and classification based on the pre-trained MATLAB function after which the mobile app outputs the diagnosis using a pre-recorded text and audio output. Thus, exploring the

possibility of employing CNN as feature learners while incorporating the benefits of standard feature reduction and classification techniques based on a mobile architecture using Unity 3D on the front end and Matlab for the pre-trained backend as shown in Figure 35.

5.1 Feature Learning

The images were first normalized to ensure uniform composition before being fed to the convolution layer. The convolution layer is divided into a convolution operation, a pooling operation, and an activation function. The convolution operation extract features by considering the values of local pixels through a matrix operation of image and filter. The Convolutional operation is expressed in equation 19, where x^i and y^j are the input maps, K^{ij} is the convolutional kernel between both maps.

$$y^{j(r)} = max\left(0, b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)}\right) \quad 19$$

The pooling operation then takes only pixel values that satisfy specific rules among the pixels within a particular region. This step quantizes the dimension of the input data which increases the processing time. This may lead to the loss of usefusl pixel values that can contribute to the recognition and detection of the medical images. Max-pooling was selected for this research using Equation 20, where each neuron in the x^i and y^j input maps pools over S x S non-overlapping local region in the ith input map x^i.

$$y_{j,k}^i = \max_{0 \le m,n < s} \{x_{j.s+m,k.s+n}^i\} \quad 20$$

The activation function is used to change the results of the previous layer nonlinearly because when linear output is employed, the learning process becomes difficult due to the disappearing gradient based on the back-propagation process. Generally, either Relu, Sigmoid, or tanh function is favored. For this research, Relu function (y = max (0, x)) is employed as it has proven better fitting capability than sigmoid and tanh function [61]. Additionally, a normalization approach termed drop out is then employed to avoid overfitting, which is a situation where more than needed data is learned from a given input thereby failing to provide adequate results for additional data. The technique reduces data size by randomly dropping nodes in the fully connected tier nodes during the learning process. Lastly, the output of the fully connected layer is used to determine the extracted feature where N is the number of features learned [113].

5.2 Feature Reduction

To further reduce the processing time while increasing the intra-feature variability, the Principal Component Analysis (PCA) was employed. It is a statistical method for reducing the extracted feature vectors of a dataset; by grouping highly correlated feature vectors such that it retains the maximum variance and remarkably minimizes the least square reconstruction error. The technique starts by normalizing the size of the dataset, computing the covariance matrix of the dataset which is then used to deduce the Eigenvectors and Eigenvalues. The Eigenvalues are sorted in descending order, and *k* Eigenvectors that correspond to the most significant Eigenvalues are chosen where k is the scale of the new feature subspace ($k \leq dk \leq d$). The projection matrix w is constructed from the selected k eigenvectors. Finally, the original data x is transformed to w using the k-dimensional feature subspace *Y* [93].

5.3 Classification

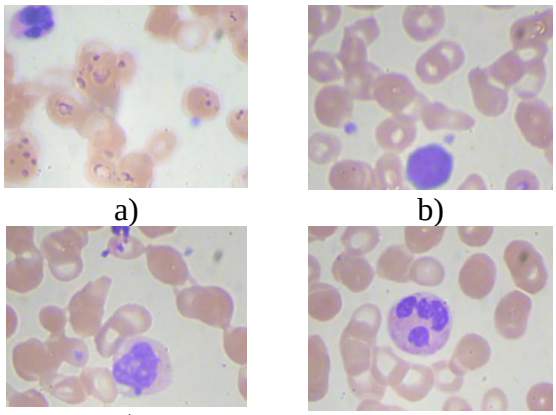
In the Classification phase of this research, ensemble classifiers are employed to combine six base classifiers to create a stronger classifier. The base learners employed are Support Vector Machine, K-Nearest Neighbor, Random Forest, Naïve Bayes, and Decision Tree. The approach combines the classification accuracy using heterogenous ensembles including stacking, bagging, and Weighted Majority Algorithm. Stacking involves the use of ensemble learners to combine the results of multiple base classifiers on a single dataset. It trains a linear regression classifier to combine the outputs of the base models.

Furthermore, the individual classifiers were trained using the whole training dataset, before linear regression is applied to fit using the accuracy and error of the base techniques in the ensemble system. Weighted Majority Algorithm (WMA) is used to construct a compound algorithm from a pool of prediction algorithms. This technique is a binary decision problem and constructs a compound algorithm by assigning a positive weight to each base classifier and computes the weighted votes of all the base models in the pool while assigning the sample to the prediction with the highest vote. It assumes no prior knowledge of the detection rate of the base algorithms in the pool but instead believes that one or more of the base classifiers will perform well [89]. Lastly, Bagging uses different subsets to train the base classifier separately, it generates multiple bootstrapped training sets by repeatedly (n times) selecting one of the n samples at random

and calls the base model learning algorithm with each of them to yield a set of base models; where each sample has an equal probability of being selected. Voting is employed to obtain the classification results and provides the ensemble with a balance between variability and similarity. Although, some training examples may not be selected at all and others may be chosen multiple times.

5.4 Data Collection

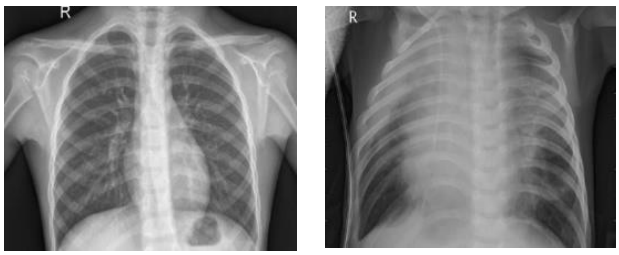
Three datasets were employed in this research; White blood cell (4 classes), Pneumonia (2 classes), and NIH (4 classes) chest x-ray images. The blood cell dataset consists of 12,500 augmented colored images of White blood cells comprising four different cell types namely Eosinophil, Lymphocyte, Monocyte, and Neutrophil approximately 3,000 images per class as shown in Figure 36 [78].



a) b) c) d)

Figure 36. Sample a) eosinophil, (b) lymphocyte, (c) monocyte, and (d) neutrophil white blood cell

Furthermore, the Pneumonia chest x-ray dataset was extracted as part of patients' routine clinical care from pediatric patients ranging from one to five years old selected from Guangzhou Women and Children's Medical Center, Guangzhou, China.



a) b)

Figure 37. The sample of a) normal, and b) pneumonia chest x-rays

Figure 37 depicts sample Pneumonia chest x-ray images, consisting of 5,863 images divided into two categories, namely Pneumonia and Normal patients [59]. Lastly, the NIH chest x-ray dataset comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. There are 15 classes (14 diseases, and one for "No findings") [54]. For this study, we employed a subset of this dataset using only four classes; Mass, Pneumonia, Cardiomegaly, and Pneumothorax comprising 5,863 NIH chest x-ray images.

5.5 Experimental Setup

In this research, a pre-trained medical image analysis model was developed using CNN and Ensemble Learners. The CNN ensemble comprises two phases; the CNN model and ensemble classifier. The CNN model includes an input, convolutional, and fully connected layer as presented in Table 15. The proposed CNN Model has three convolutional layers ranging from convolutional layer 1 to 3. The feature extraction of the medical image is conducted on the three convolutional layers; it uses several filters with varying dimensions containing two max-pooling and batch normalization layers. The Rectified linear unit (Relu) layers were implemented in all the Convolutional Network; which acts as a threshold function to reset all negative input values x to zero. This, in turn, increases the generalization, reduces computation time, and learning the speed of the system. The learned feature output is extracted by using the fully connected layers based on nodes constructed at each layer, where the number of nodes in the layer is equal to the number of features extracted using CNN. The Medical Image Analysis system was pretrained using Unity 3D and MATLAB 2018a on a Windows 10 Pro operating system by implementing CNN on a

computer with an Intel® Core™i7-6700 CPU at 3.40 GHz, 64 GB memory, and an NVIDIA GeForce GTX TITAN X graphics card with 3072 CUDA cores, and 16 GB GDDR5 memory before its deployment on Samsung S8 for validation and testing. We employed Cross-Validation for the analysis of the database. It is a statistical technique utilized in evaluating predictive models by dividing the data into training and validation sets. The method randomly separated the dataset into K-fold equal data sizes and crossed over each K-dataset partition in successive rounds such that each data point has a chance of being in the validation process. The proposed method employed a 10-fold Cross Validation tested on a randomized sample of each dataset. The actual learning of the techniques was carried-out on the training set, while the validation set was used to finetune the hyper-parameters and the overall performance of each system was evaluated on the test set. As principal evaluation measure, we plotted the ROC curve (sensitivity by 1-specificity), training time, objective, and top 1 error of the different classes. Due to the increased sensitivity to imbalances among the classes, the overall system accuracy was also computed.

Table 15. CNN model at each stage of the convolutional and fully connected layers

Layer Type	**0 input**	**1 Conv1**	**2 bnorm1**	**3 relu1**	**4 pool1**	**5 Conv2**	**6 bnorm2**	**7 relu2**	**8 pool2**	**9 Conv3**	**10 relu3**	**11 Fcn1**
Filter Dimension	n/a	3	n/a	n/a	n/a	15	n/a	n/a	n/a	20	n/a	100
Filter Dilation	n/a	1	n/a	n/a	n/a	1	n/a	n/a	n/a	1	n/a	1
Number of Filters	n/a	15	n/a	n/a	n/a	20	n/a	n/a	n/a	100	n/a	4
Stride	n/a	1	1	1	2	1	1	1	2	1	1	1
Rf size	n/a	7	7	7	8	12	12	12	64	64	64	64
Rf offset	n/a	4	4	4	4.5	6.5	6.5	6.5	32.5	32.5	32.5	32.5
Rf stride	n/a	1	1	1	2	2	2	2	54	54	54	54
Data size	64	58	58	58	29	27	27	27	1	1	1	1
Data Depth	3	15	15	15	15	20	20	20	20	100	100	4
Number of Data	256	256	256	256	256	256	256	256	256	256	256	256

5.6 Result and Evaluation

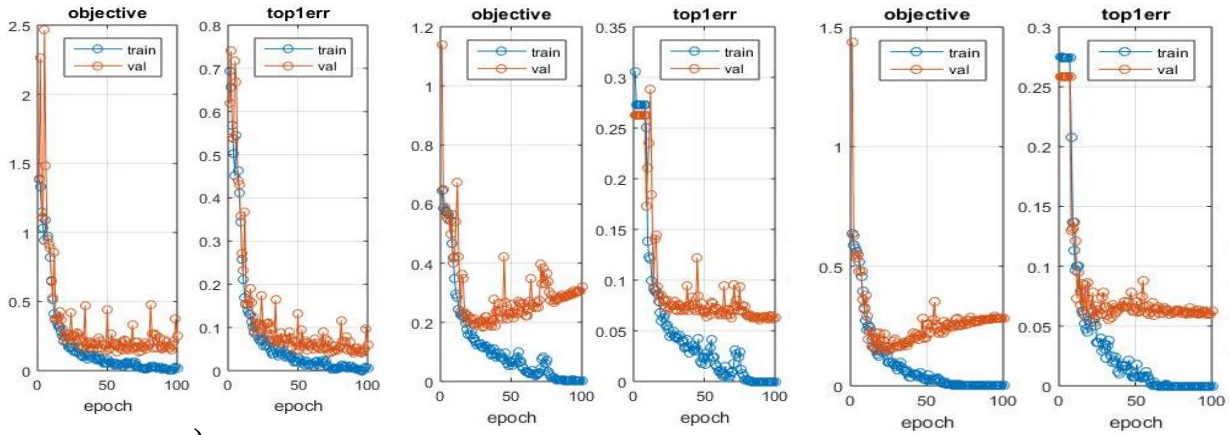


Figure 38. Objective and top 1 errors of all dataset for the proposed system (CNN Bagging)

Figure 38 presents the objective and top 1 errors of the proposed system using a) White blood cell, b) Pneumonia and c) NIH chest x-ray dataset. It evaluates the training loss (blue) and validation loss (dashed orange) across training epochs passed over the entire training set of images broken up into batches. The graph shows how often the highest scoring guess is wrong using both the training and validation set. The result of the system showed that the training and validation loss reduced monotonically and reached stability at 75 epochs with the chest x-ray database. However, it took longer to converge for the White blood cell dataset which reaches stability at 100 epochs as the backpropagation updates to the weights of the network in a direction that should decrease the loss or slightly increased the loss.

Table 16. Accuracy (%) of white blood cell diagnosis using several CNN based methods

White Blood Cells	25 Epochs	50 Epochs	75 Epochs	100 Epochs
CNN	91.89	93.73	94.11	94.71
CNN+SVM	85.70	88.50	90.39	90.39
CNN+KNN	86.47	88.09	89.40	89.40
CNN+RF	83.35	85.81	88.14	88.06
CNN+NB	83.87	85.24	87.57	87.57
CNN+DT	78.77	83.19	85.35	89.18
CNN+WMA	86.22	88.44	89.95	89.98
CNN+Stacking	86.36	88.46	90.24	90.26
CNN+Bagging	87.13	89.39	90.46	91.37

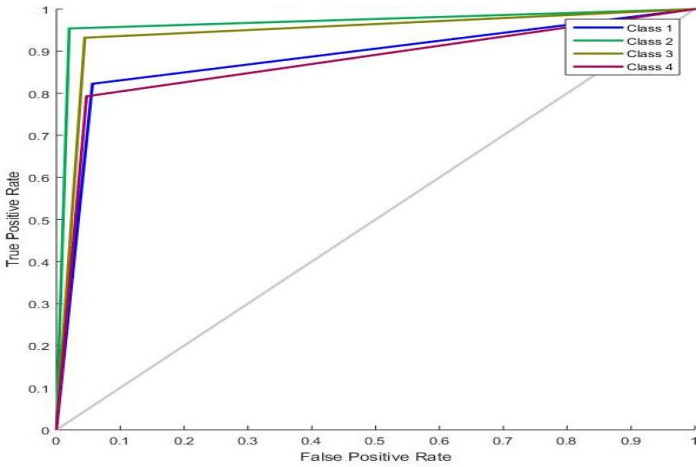
Table 17. Accuracy (%) of pneumonia NIH chest x-ray diagnosis using several CNN methods

Pneumonia chest x-ray dataset	25 Epochs	50 Epochs	75 Epochs	100 Epochs

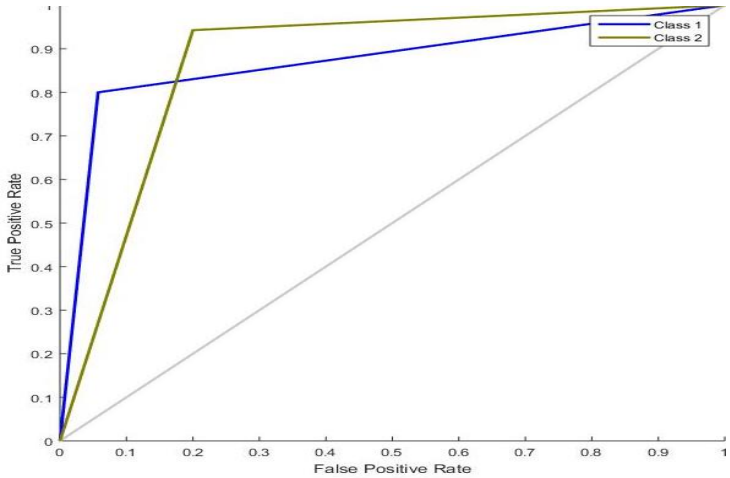
CNN	90.52	93.43	94.29	94.29
CNN+SVM	91.21	91.21	91.21	91.21
CNN+KNN	90.41	90.41	90.41	90.41
CNN+RF	89.66	89.89	89.78	89.83
CNN+NB	88.81	80.81	88.81	88.81
CNN+DT	88.86	88.86	88.86	88.86
CNN+WMA	90.52	90.58	90.52	90.52
CNN+Stacking	91.09	91.44	91.49	91.32
CNN+Bagging	91.23	91.77	91.97	90.83

Table 18. Accuracy (%) of NIH chest x-ray diagnosis using several CNN based methods

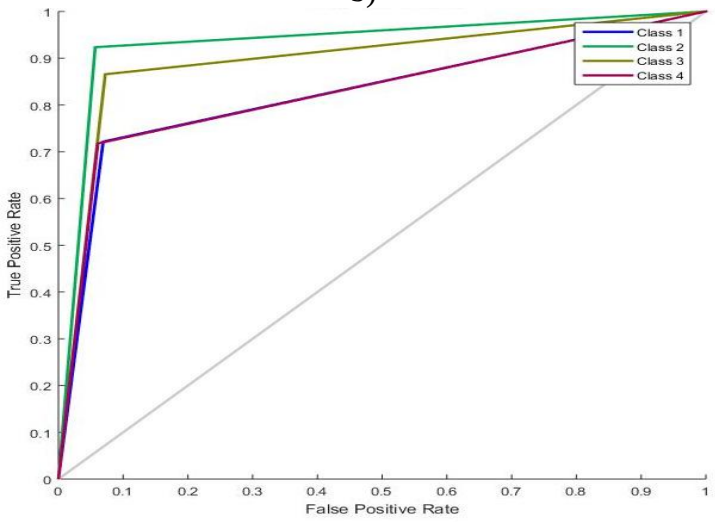
NIH chest x-ray dataset	**25 Epochs**	**50 Epochs**	**75 Epochs**	**100 Epochs**
CNN	74.94	77.84	79.89	78.29
CNN+SVM	72.72	75.83	77.10	76.85
CNN+KNN	72.03	75.40	76.74	76.22
CNN+RF	70.66	74.71	74.83	74.94
CNN+NB	70.53	72.61	74.85	73.24
CNN+DT	70.01	71.15	70.55	71.07
CNN+WMA	72.62	75.79	77.05	76.74
CNN+Stacking	72.77	75.88	77.38	76.84
CNN+Bagging	72.94	76.09	77.63	76.98



a)



b)



c)

Figure 39. ROC curve of a) white blood cell, b) pneumonia, and c) NIH chest x-ray images

Tables 16, 17, and 18 present the classification accuracy using extracted CNN features diagnosed by classifiers and ensemble learners for White blood cell, Pneumonia, and NIH chest x-ray dataset. The accuracy of the White blood cell dataset converged at 100 epochs while both chest x-ray set converged faster at 75 epochs across all techniques. It shows how the optimized feature set using CNN when classified by Bagging achieve better accuracies. Although the accuracies for CNN achieved slightly better (about 1%) than the CNN ensemble feature set, we have a feature reduction ratio by half of the features using PCA that reflects on time and resource consumption. It can be deduced that the approach is useful in medical image analysis. It also verifies that the CNN descriptor is a viable method and there is no significant difference using colored (White blood cell) and grayscale (Pneumonia and NIH chest x-ray) image format. The accuracy of the CNN proposed feature extraction method using different learning methods was also measured. Although their accuracy was very promising, the proposed approach feature set when classified using Bagging achieved the highest accuracy for both datasets. It can be deduced that our methods performed better with colored images than grayscale images as Pneumonia with two class database provided competitive accuracy and the White blood cell set having four classes but presented comparably lower results like the NIH chest x-ray dataset with four classes.

Table 19 presents the average processing (training and testing) time for the White blood cell, and chest x-ray set employed using CNN, other classifiers and the ensemble method. To improve the recognition time, each system was pre-trained before testing on a mobile phone. The time for each set computed with different classifier as each technique varied because of the architecture's approach. It shows the training and testing time for the dataset with each method, the grayscale image (Pneumonia and NIH chest x-ray) had the shortest processing time for CNN+KNN, CNN+DT, CNN while the colored image (White blood cell) presented the recognition time with the CNN+SVM, CNN+RF, CNN+NB, CNN+WMA, CNN+Stacking, and CNN+Bagging. For the training phase, the CNN method had the shortest time with the proposed classification method while the CNN+Bagging had the highest processing time. During the testing stage, CNN+KNN method had the shortest time with the proposed classification method with CNN giving the highest processing time. The feature extraction and the feature reduction processing were taken into consideration as it contributed to the processing time. However, the validation time differences for each classifier were small regardless of the dataset. Although all other classifiers have a remarkable computation time, the proposed CNN needed less than 2 seconds to classify all the medical images.

We realized that the image format either colored or grayscale did not pose a significant extension to the training or testing time. However, embedding the MATLAB backend into the mobile app architecture reduced the testing time for all methods except for CNN due to the SoftMax computation time.

Table 19. Training and testing time (sec) for the different CNN based methods

Methods	White Blood Cells		chest x-ray	
	Training Time (s)	Test Time (s)	Training Time (s)	Test Time (s)
CNN	1.86	1.71	1.04	0.71
CNN+SVM	1.74	1.18	1.34	0.52
CNN+KNN	1.01	1.01	1.40	0.49
CNN+RF	2.59	1.20	1.53	0.23
CNN+NB	2.42	1.25	1.47	0.30
CNN+DT	2.21	1.32	1.69	0.34
CNN+WMA	2.08	1.64	1.85	0.38
CNN+Stacking	2.72	1.66	1.94	0.66
CNN+Bagging	2.65	1.63	1.86	0.64

For a more detailed comparison at different operating points, we also performed a receiver operating characteristic (ROC) analysis for the proposed system using a) the White blood cell b) Pneumonia and c) NIH chest x-ray classes as presented in Figure 39. The average ROC curves over the different classes were plotted to determine how effective our systems can distinguish medical image samples and predict the classes. The results of the analysis confirmed the statistically superior performance of the proposed CNN Ensemble against all other methods experimented in this research. Further experiments were conducted in Table 20 using the accuracy of traditional feature extraction techniques on existing classifiers in addition to the proposed ensemble approach. The table presents the recognition rate of the different descriptors and classifiers on White blood cell and chest x-ray dataset; each feature extraction approach was applied to all the descriptors to determine the best pair with the highest recognition rate. The proposed approach CNN+Bagging presented the best recognition rate after Classification while SC+KNN had the worse accuracy for White blood cell and chest x-ray dataset respectively. The rankings of the descriptor and classification technique ranging from worse to best are SC, SIFT, HOG, SURF, CNN and KNN, NB, RF+DT, WMA, SVM, Stacking, Bagging respectively. From the result, SC had the lowest acceptance rate across all datasets while CNN presented the best results across all dataset with relatively higher performance than HOG, SIFT, and SURF. Regarding dataset, White blood cell had the best result while chest x-ray had the worse rate

although this could be attributed to the set image channel. Lastly, the proposed approach presented promising results in comparison with existing techniques and can be engaged in descriptor feature extraction.

Table 20. Accuracy (%) of feature learners with bagging on different classifiers

Set	Feature	SURF	SIFT	HOG	SC	CNN
White blood cell	KNN	70.73	65.27	71.90	63.45	89.40
	RF	75.28	71.93	75.62	71.44	88.06
	NB	70.00	72.24	74.99	79.45	87.57
	DT	76.51	75.27	75.89	73.89	89.18
	SVM	77.39	76.29	76.21	75.11	90.39
	WMA	76.74	74.38	76.46	73.33	89.98
	Stacking	77.53	75.71	76.48	74.72	90.26
	Bagging	85.40	84.56	87.60	82.45	91.37
chest x-ray	KNN	68.00	65.89	67.51	61.47	90.41
	RF	73.11	71.64	73.07	69.38	89.83
	NB	71.91	70.95	71.74	66.38	88.81
	DT	73.11	72.89	73.55	70.63	88.86
	SVM	77.38	76.32	76.21	73.70	91.21
	WMA	76.18	75.47	76.05	77.23	90.52
	Stacking	80.92	76.54	78.21	78.93	91.49
	Bagging	83.46	82.95	85.53	81.56	91.97

Table 21 displays the precision and processing time of White blood cell and chest x-ray dataset in the medical image analyzer using Shape Context (SC), Scale Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF), Histogram of Gradient (HOG), and CNN using our proposed bagging classifier. This was experimented to evaluate the handcrafted feature extraction against the CNN model using our proposed Ensemble Classifier to determine which of the techniques can best predict the image class. From the result, the proposed CNN model technique presented the best feature vector results for this experiment while SC had the lowest rate across all databases. NIH chest x-ray images had the worse result across all methods, and White blood cell gave the best precision. The processing image time for the datasets differed broadly across methods, and the chest x-ray had the shortest processing time, and the White blood cell had the longest time. Each technique extracted a different number of features hence the disparity in timing. Additionally, the training time for all dataset was relatively close. The testing time for all dataset was also in the same range although CNN had longer training time from all the handcrafted feature techniques.

Table 21. Precision (%), training and testing time (sec) using feature learners with Bagging.

	White blood cell			chest x-ray		
	Precision	**Train Time**	**Test Time**	**Precision**	**Train Time**	**Test Time**
SURF	75.40	2.12	1.37	73.46	1.35	0.44
SIFT	74.56	2.68	1.21	72.95	1.64	0.59
HOG	77.60	1.94	1.43	75.53	1.72	0.56
SC	72.45	2.41	1.59	71.56	1.29	0.28
CNN	91.37	2.65	1.63	91.97	1.86	0.64

Table 22. Comparison of the proposed CNN bagging with published state-of-the-art methods

Recognition Rate	**Technique**	**White blood cell**	**chest x-ray**
Proposed Method	CNN Bagging	91.37	91.97
Handcrafted feature method	Theera-Umpon & Dhompongsa 2007 [115]	83.29	N/A
	Habibzadeh et al. 2013 [46]	84.77	82.91
	Su et al. 2014 [111]	89.72	88.89
	Prinyakupt & Pluempitiwiriyawej, 2015 [90]	79.25	78.78
Existing Deep Networks	Li et al. 2017 [66]	97.47	96.84
	Salehinejad et al. 2017 [104]	100	100
	Ypsilantis & Montana 2017 [133]	99.46	98.00
	Brestel et al. 2018 [15]	100	100

Table 22 shows a comparison of the performance of the proposed CNN Ensemble with different state-of-the-art methods both handcrafted and deep networks to validate the efficiency and accuracy of our proposed system. The handcrafted features employed included wavelet transform, Histogram of Gradient, Geometry, color, binary, statistical and nucleus-based feature while the deep network used were DCNN, RNN, ResNet, and RadBot-CXR architectures. The suggested deep learning-based algorithm has improved the detection accuracy by more than 10% on average compared to previous feature-based detection algorithm. For some feature extraction algorithm, the proposed algorithm shows a significant difference of more than 15% outperforming handcrafted techniques. The deep neural network presented better recognition rate than the proposed system it, however, required more memory space and training time, making our approach more valid for mobile-based application also inferring the usefulness of the procedure for medical image analysis. Regarding the image format, the colored image (White blood cell) presented significant impact on the training and testing time in comparison with the grayscale images (Pneumonia and NIH chest x-ray); yet the precision was higher with the colored image. Lastly, the

system administrator can edit the input text which automatically updates the audio as the App utilizes a text to audio converter that gives an audio description of the currency being tested by the mobile application.

Chapter 6

6 Semi-supervised Learning with Bidirectional GAN Using Capsule Discriminator

Generative Adversarial Network is a generative model with the widespread success that can be used to generate photo-realistic images. It is a two-player min-max game between a generator *G* and a discriminator *D*. The generator employs noisy data to produce possible sample images to deceive the discriminator while the discriminator tends to improve its discriminative ability by learning [33].

6.1 BIGAN Architecture

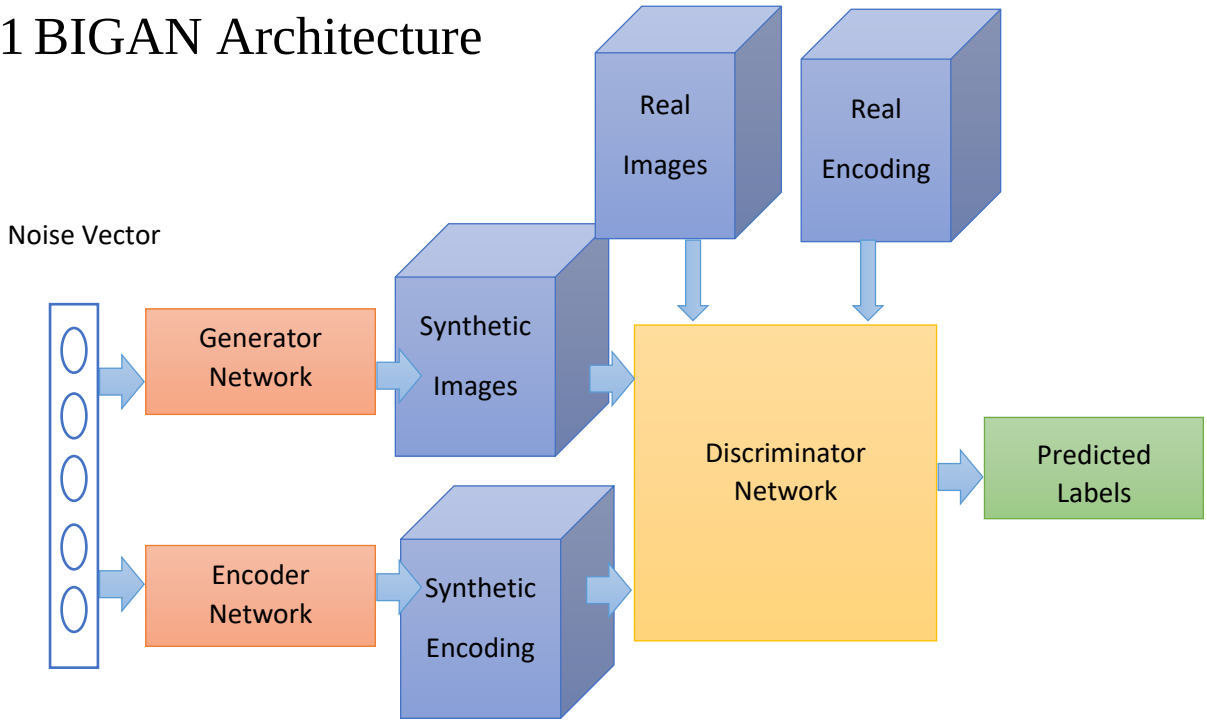


Figure 40. The BIGAN architecture

We propose an extension to the Bidirectional Generative Adversarial Network (BIGAN) architecture that learns the inverse mapping to project the data back into the latent space. It has a

perceptual loss function "The Discriminator" using Binary Cross-Entropy (BCE) as its loss function considers perceptual information. BIGAN trains not only a generator but also an encoder that induces a distribution for mapping data point into the latent feature space of the generative model. Furthermore, the discriminator *D* discriminates not only in data space (*x* versus *G(z)*), but jointly in data and latent space (tuples (*x*, *E(x)*) versus (*G(z)*, *z*)), where the latent component is either an encoder output E(x) or a generator input z. The discriminator is modified to take input from the latent space, predicting *PD* (*Y* |*x*, *z*), where *Y* = 1 if x is real (sampled from the actual data distribution *px*), and *Y* = 0 if x is generated (the output of *G(z)*, $z \sim pZ$). The generated image together with the latent input noise (*G*(*z*), *z*), and the real data along with its encoded latent representation (*x*, *E*(*x*)) are fed into the discriminator *D* for classification [29]. The structure of the proposed system is depicted in Figure 40. The BIGAN training is defined as a min-max objective:

$$min_{G,E} max_D \; BIGAN\;(D,E,G) \tag{21}$$

Where

$$BIGAN\;(D,E,G) = E_{x\sim px}\frac{[E_{z\sim PE(.|X)}[\log D\;(x,z)]]}{\log D(x,E(x))} + E_{x\sim pz}\frac{[E_{x\sim PG(.|Z)}[\log(1-D\;(x,z))]]}{\log(1-D(G(z),z))} \tag{22}$$

6.2 Generator

Figure 41. Structure of the generator (DCGAN Architecture)

A class of CNN called deep convolutional generative adversarial network was employed as the generator. The approach modified some critical areas of the CNN architecture. The first was replacing max-pooling with strided convolutions, allowing the network to learn its spatial downsampling. The second was eliminating fully connected layers on top of convolutional features. The third is including batch normalization which stabilizes learning by normalizing the input but not applying to the output layer. Finally, the Relu activation is employed in all layers of

the generator except for the output which uses a Tanh function. [91]. As shown in Figure 41, the DCGAN generator included four convolutional layers with a Relu activation and batch normalization for in all layers except for the output which generally utilizes a Tanh activation. We also employed a Reshape function in the first layer to change the dimensions of the input, without changing its data and two Upsampling functions in the first two layers to improve the image resolution.

6.3 Encoder

While the Encoder applied the DCGAN approach; it takes as an encoding blurred with a Gaussian noise patch and transforms it into the data form we are interested in imitating, in our case an NIH chest x-ray encoding. The Encoder consists of three layers. The first layer had a flatten function to restructure the input dimension; then the dense and batch norm layers were implemented in all the Convolutional Network, and each layer employed a Leaky ReLU except the output layers that comprised Relu instead of the recommended Tanh for DCGAN as depicted in figure 42. This network architecture was selected to generate new samples from an unknown probability distribution due to its fast convergence capability. The capacity of the generator model to efficiently serve as a density model of the training data, and its efficient and straightforward sampling [91].

Figure 42. Structure of the encoder (DCGAN Architecture)

6.4 Discriminator

We extended the Capsule network by first starting with a Capsule that includes convolutional and batch normalization. The next layer employed is a typical primary Capsule Comprising Convolution, Reshape and Squash function. The input image is fed into a convolution layer that uses a filter to output an array of feature maps. We apply the Reshaping function to these feature maps and reshape into the 8D vector for all feature maps combined as a collection of activations

representing the orientation of the image and the intensity of the vector which denote the presence of the image. The last function used is the squash function as shown in equation 23; where *Vj* is the output the vector of Capsule *j* and *Sj* is its total input.

$$v_j = \frac{||s_j||^2}{1+||s_j||^2}\frac{s_j}{||s_j||} \quad 23$$

It outputs a value between 0 and 1 for each Capsule where the length of the input decides if the object is in that given location in the image. The next step is determining how the Capsule (*i*) in a lower-level layer will send its output vector to higher-level Capsules (*j*). This decision is made by changing the scalar weight C_{ij} =coupling coefficient that multiplies its output vector before being treated as an input to a higher-level Capsule as shown in equation 24. Where the total input to a Capsule Sj is a weighted sum over all prediction vectors $\hat{U}_{j|i}$ from the Capsules in the previous layer [128].

$$s_{j=\sum_i c_{ij}\hat{U}_{j|i}}, \ \hat{U}_{j|i} = W_{ij}u_i \quad 24$$

The coupling coefficients between the Capsule (*i*) and all the Capsules in the previous layer sum to 1. We treat the coupling coefficient as a SoftMax over bias weights from the previous dense layer. The SoftMax make sure that each weight c_{ij} is a non-negative number and their sum is equal to one. Whose initial logits b_{ij} are the prior log probabilities that Capsule *i* should be coupled to Capsule *j* that is achieved using equation 25.

$$c_{ij} = \frac{exp(b_{ij})}{\sum_k exp(b_{ik})} \quad 25$$

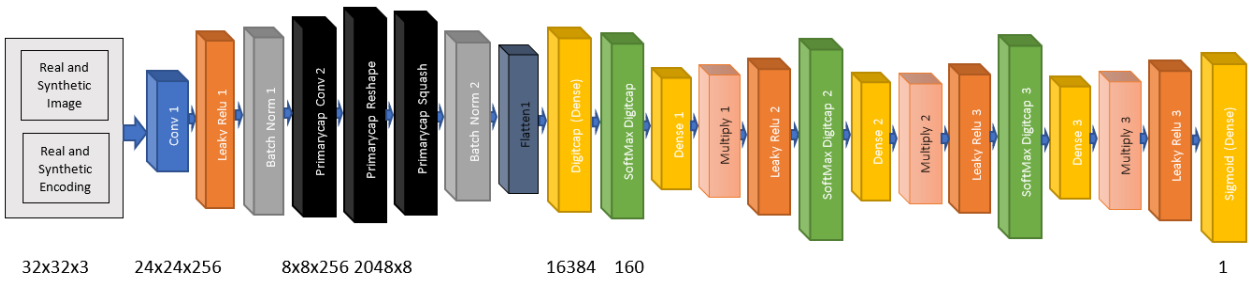


Figure 43. Structure of the discriminator (Capsule Network)

Instead of Squashing the Capsule outputs we replaced it with Leaky Relu as Squashing function created severely blurry artifacts. We repeated the routing agreement two more times with Leaky Relu, and the loss of the output of the final layer was trained using a sigmoid cross-entropy as shown in Figure 43 [101].

6.5 Data Collection

Three datasets were employed in this research; CelebA, Pneumonia, and NIH chest x-ray images. The CelebA included about 202,599 face images in batches of 20 images from approximately 10,000 celebrities; it provided a set of pre-cropped face images aligned using the hand-labeled key points [70]. The Pneumonia chest x-ray set consisted of 5,863 images grouped into two classes, Pneumonia, and Normal patients. The images were extracted from the pediatric patient as part of patients' routine clinical care ranging from one year to five years at Guangzhou Women and Children's Medical Center in Guangzhou, China [59]. Lastly, the NIH chest x-ray dataset contains 112,120 chest x-ray images having disease labels from 30,805 unique patients. There are 15 classes: Atelectasis, Fibrosis, Effusion, Infiltration, Pneumothorax, Edema, Pneumonia, Pleural-thickening, Cardiomegaly, Nodule Mass, Consolidation, Emphysema, Hernia and one for No findings [54].

6.6 Experimental Setup

The evaluation of the images was conducted using the training, and testing approach based on the semi-supervised learning scheme. The data was first learned with the hyper-parameters fine-tuned on the training set, and the overall performance of the system was assessed on the test set. The principal evaluation measures employed include plotting the generator loss, and discriminator loss and accuracy. We also computed the best latent space for the dataset with the precision, recall, inception score and Mean Square Error for the system. The setup for our experiments is as follows:

Hardware: Tesla K80, having 2496 CUDA cores, 12GB GDDR5 VRAM, 1 x single core hyperthreaded Xeon Processors @2.3Ghz (One core, two threads), 12.6 GB memory, and 33 GB SSD.

Software: Google Colab Jupyter notebook is running on Google Cloud Compute Engine: Ubuntu Linux 17.10 OS, containing all the configuration needed for this experiment, such as Keras, TensorFlow, and Python.

Test parameters: kernel size: 28x28 and 32x32, Adam optimizer, beta1 (momentum of Adam optimizer) = 0.5, learning rate = 0.0002. Mini-batch sizes range from 1, 32, to 256 (i.e., each epoch contains only one batch).

Training: The system feeds the images to the encoder while feeding the noise and input data to the generator to train them. Furthermore, the semi-supervised approach learns by selecting a random half batch of images, generating half batch of new images before picking a random batch to encode and train the discriminator by feeding both real and fake images one by one into the BIGAN system. The approach only learns the generator after combining the encoder and generator to distinguish generated images.

6.7 Result and Evaluation

The discriminator outputs 1 when it predicts an image as real, and 0 is detected as fake images. So, the discriminator loss is generated by computing the sum of the loss of the fake images and the real images. The loss of the real and fake image is derived by comparing the prediction to the image labels to determine if it is correct. While for the generator loss, we do the same as in the last step, but instead we check the prediction to confirm if we were able to fool the discriminator.

Figure 44 presents the generator loss while training on a) CelebA b) Pneumonia and c) NIH chest x-ray dataset. In the first case, the loss of the model is bouncing around throughout all 30000 epochs although it naturally decreases over time from 28.5 and flattens at 2 for CelebA, 43.5 to 7 for the Pneumonia while the NIH chest x-ray images reduced over time from 16 and flattened at 1. It is quite easy to indicate the flattening since it fluctuated only a few more time. However, the pattern does not seem to leave the boundaries of 45 to 5 for both CelebA and Pneumonia and 16 to 1 for the NIH chest x-ray dataset that we consider as flattening. We assume those fluctuations stem from the model trying to improve itself. It is necessary to note that the loss numbers themselves are usually non-informative. The loss flattening would typically signify that the model found some optimum and it cannot improve more, which actually means that it has learned to map the noise vector to a Z representation that is enough to challenge the discriminator. While these observations are valid for both CelebA and Pneumonia set, the NIH chest x-ray set faces fewer fluctuations and goes down to convergence somewhat faster than the CelebA and Pneumonia example. It is hard to judge the convergence level based on flattening and fluctuations although

the difference between losses of the first and second plots are apparent, but the loss in CelebA and Pneumonia plot might be staying low and feeding the discriminator with meaningless outputs without necessarily having a mode collapse. GANs losses are very non-intuitive since generator and discriminator are competing against each other. Hence improvement on the one means the higher loss on the other.

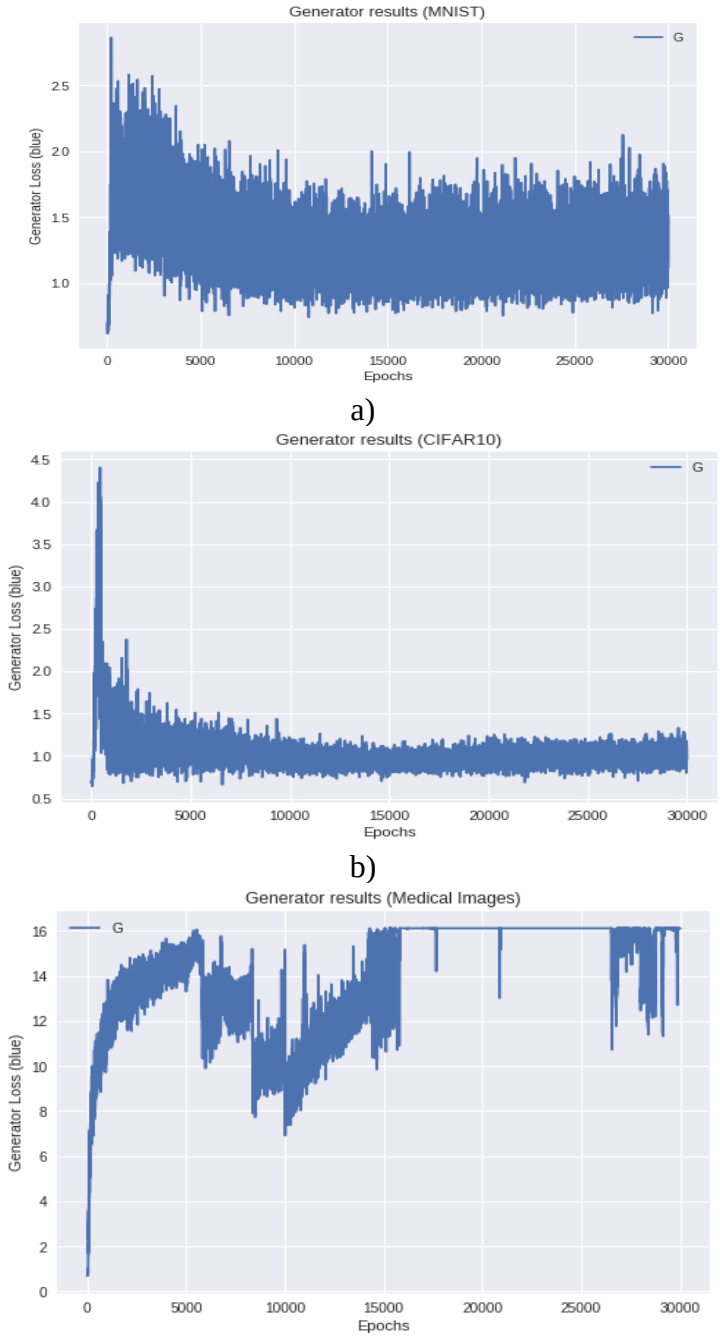
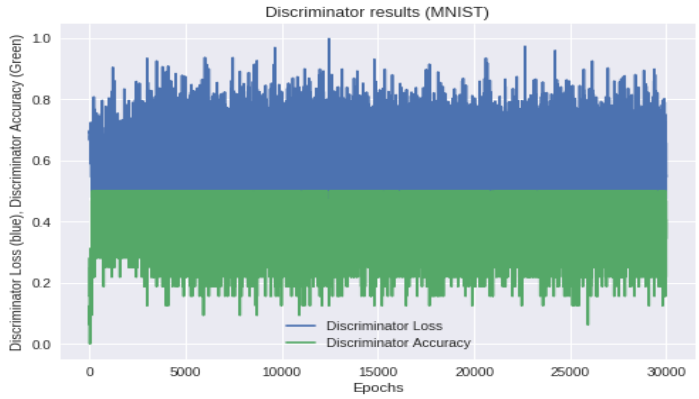
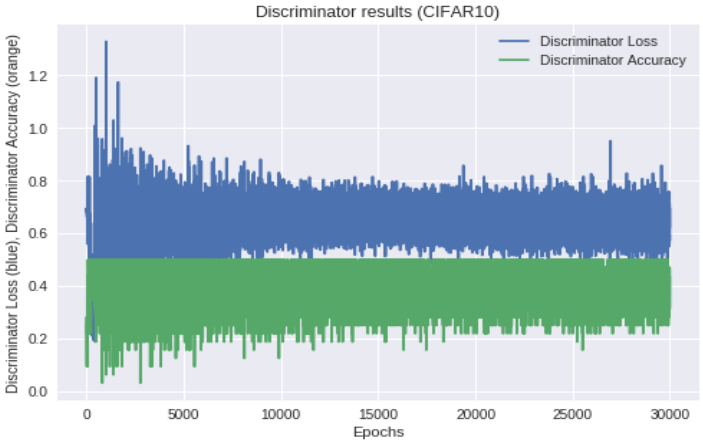


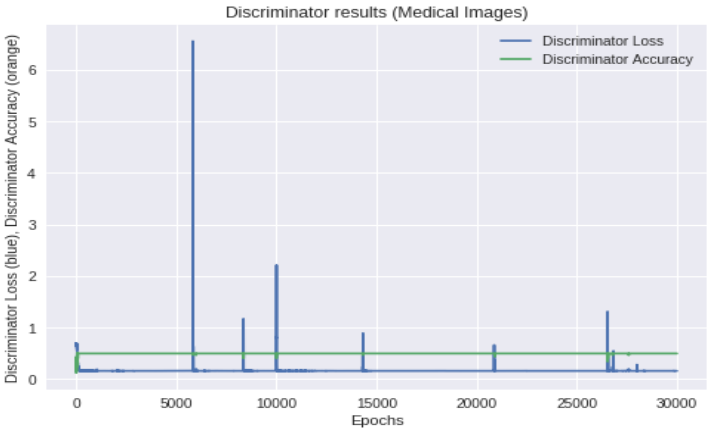
Figure 44. Generator losses for a) celeba, b) pneumonia, and c) NIH chest x-ray dataset



a)



b)



c)

Figure 45. Discriminator losses/accuracy for a) celeba, b) pneumonia, and c) NIH chest x-ray dataset

Figure 45 shows differences between the discriminator loss and accuracy while training on a) CelebA b) Pneumonia and c) NIH chest x-ray dataset. Accuracy is plotted upturned relatively to the loss because its values do not reflect the actual progress for the CelebA and Pneumonia. For the CelebA and Pneumonia; it started with a high loss rather than a relatively low loss noticed in

the chest x-ray sets, although the method had a similar fluctuating pattern, the loss slightly reduced from 10 to between 9 to 7 for CelebA and from 13 to between 12 to 7.5 for Pneumonia. Although this loss had a more usual pattern proving a correlation between the loss of the discriminator and the loss of the generator both dropping suddenly at 9000, 18000 and 2500 epochs signifying a very minutia partial mode collapse pattern which is known as a failure of GANs. While for the NIH chest x-ray the discriminator loss had a much trickier pattern, the loss reached its highest point at about 6000 epoch reaching stability at 0.2 with the maximum loss of 6.8.

Although the loss had a more usual pattern a correlation, it can be deduced between discriminator and generator loss both converging suddenly at 10000, 15000 and 27000 epochs signifying a very minutia partial mode collapse pattern which is known as a failure of GANs. In the same view, the accuracies of all the dataset did tell us much since there is no clear optimum in practice despite research that is done on that topic nowadays. The accuracy value of the discriminator directly cannot be mapped to the output image quality. However, the accuracy scores improved ranging around 70% to approximately 95% throughout the training process for the three datasets.

Table 23. Inception score and mean square error on NIH chest x-ray dataset

Techniques	Inception Score	Mean Square Error
Extended BIGAN	13.92	0.91
BIGAN	13.58	0.92
AAE	13.11	0.95
DCGAN	10.24	1.68
LSGAN	12.33	1.42

Furthermore, we compared the extended BIGAN with other state-of-the-art generative models on the NIH chest x-ray dataset in Table 23. The visual quality of generated images is measured by the widely used metric of Inception score and Mean Square Error. The Inception Score (IS) is one of the relatively stable parameters that has been employed for assessing the quality of images generated using GANs. The metric aims to automatically evaluate the quality and robustness of the produced samples. They were shown to relate well to the human scoring of the authenticity of generated images from all dataset. While the Mean Square Error was computed, it is the sum of squared distances between our target variable and predicted values. The results of semi-supervised classification experiments on the NIH chest x-ray dataset are reported. The architecture of the semi-supervised system had dataset with 500 labels, the Inception Score of extended BIGAN is

significantly better than DCGAN and LSGAN, but on par with the original BIGAN and AAE. We also computed the Mean Square Error on the semi-supervised Extended BIGAN model and in comparison, with existing methods performed better with an MSE of 0.91. Furthermore, the BIGAN network with the same architecture produced the closest result with an MSE of 0.92, while AAE, DCGAN, and LSGAN had quite higher results in contrast with BIGAN based methods, which highlights the bidirectional architecture of the process in mitigating the effect of the BCE loss using the perceptual loss function of the discriminator.

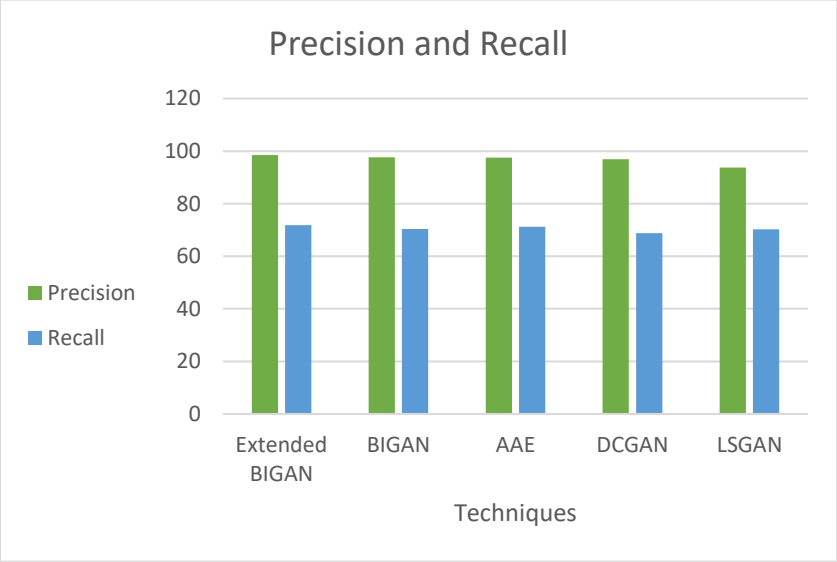


Figure 46. Precision and recall (%) on the NIH chest x-ray dataset based on the discriminator output

We also computed the precision and recall for the chest x-ray set using the discriminator results as shown in Figure 46. The extended BIGAN obtains slightly higher precision and recall than BIGAN, yet the extended BIGAN also generates very high-quality samples and outperforms all previous methods regarding the discriminator results (Capsule Network). BIGAN, either with or without the extension, achieved significantly higher results and better sample quality than the baseline DCGANs, LSGAN and AAE. Thus, demonstrating the effectiveness of the proposed approach. Although the results correlate well with the visual quality of the images, it seems to still not wholly cater to diversity issues. Furthermore, all dataset was trained using the same system setup, CelebA dataset results were able to reproduce distinguishable results already from ~5000th epoch while Pneumonia dataset reproduces satisfactory results satisfactory after the ~7500th epoch. NIH chest x-ray dataset achieves discernible results starting from ~13000th epoch. The best latent space size for each dataset employed was 100. Furthermore, we were able to address the

representational inefficiencies of CNN with transformation vectors, allowing networks to learn part or whole relationships automatically and thus generalize to novel viewpoints, an area traditional CNN's performed poorly as GAN discriminators.

DCGAN was chosen for the generator and encoder architecture as it was yet to be leveraged for the semi-supervised task while Capsule-layers was utilized in discriminator instead of a convolutional layer as it fundamentally performs a two-class classification task. Also, Capsule network was not chosen for as the generator technique because the model strength can start harshly penalizing the generator early in the training process, which will cause the generator to either fail or suffer from mode collapse.

Chapter 7

7 Improving Medical Image Resolution Using GANs

This research proposes an extension to SRCNN by updating the filter size and feature map of the low and high-resolution image. It then breaks up the images into patches for training to reduce memory consumption and handling big images of the medical dataset. The approach applied some modification to the SRGAN to handle grayscale medical dataset by stacking the input on the three layers as if it was an RGB before converting the output to grayscale. In addition to determining the best layer in the VGG19 network to compute the loss function by selecting the best patch size in the discriminator. Lastly, combining the SRGAN, using VGG19 network for feature extraction, setting discriminator network's working space as feature space, and adding the loss function based on the mean square error of pixel space, then gain more details by incorporating SRCNN layers to increase the PSNR in the reconstruction at the same time as shown in Figure 47.

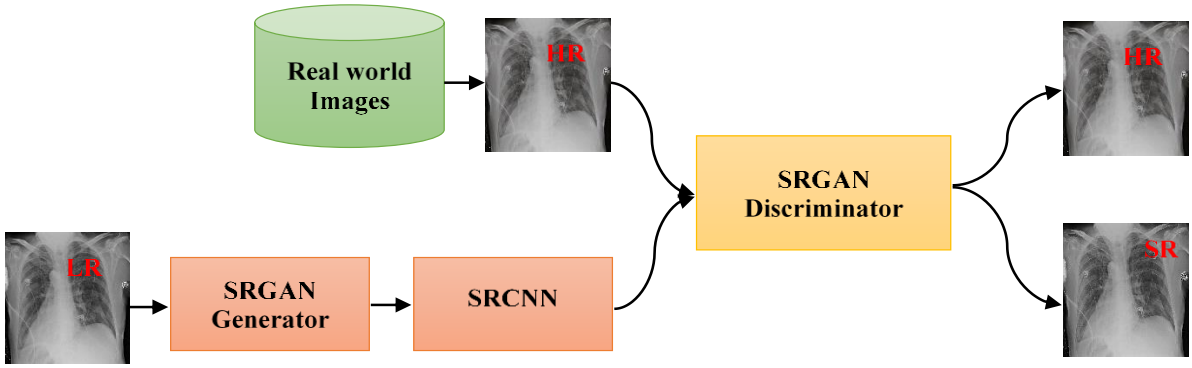


Figure 47. The architecture of the SRCNN-GAN architecture

7.1 SRGAN

Figure 48 shows the generator architecture consisting of an input layer which is convolutional and takes the low-resolution image with a scale of 0.25 to the real image (Ground truth). It also included 19 residual blocks with skip connections to overcome the degrading problem with deeper layers. We added both the output of the residual blocks with the output of the first convolutional

layer to keep the low feature information such as edges. Lastly, it included a final layer which is a deconvolutional layer to produce our output image. While the discriminator comprises of Input layer that includes both high-resolution image (produced from a generator or a real image) and validity which indicates if it is fake or real. It also comprises of 10 blocks of convolutional layers and a Final dense layer to predict if it is fake or real as shown in the discriminator architecture in Figure 49. Both showing corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

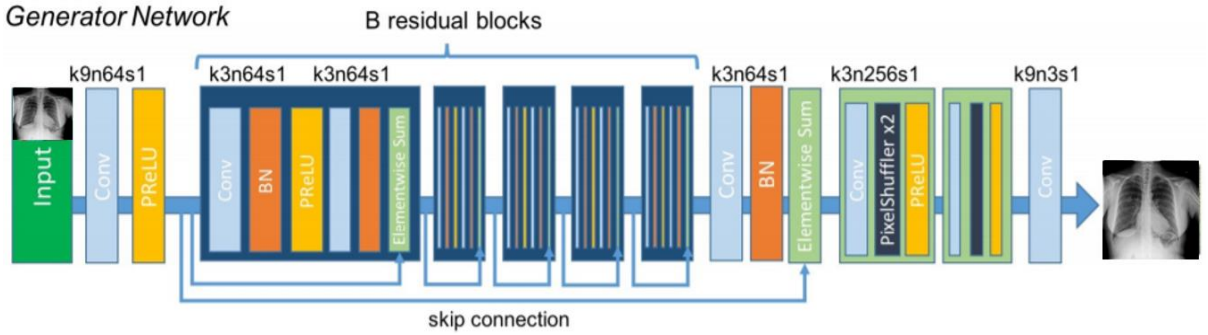


Figure 48. SRGAN generator architecture

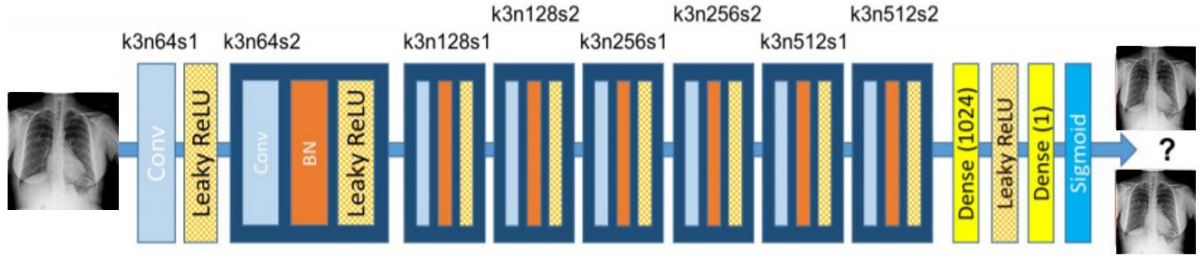


Figure 49. SRGAN discriminator architecture

We applied some modifications to the SRGAN Architecture in order to work with the medical dataset. The VGG network was initially trained on colored images, so we attempted to fine-tune the VGG network on the grayscale dataset. However, the generator result was bad, we separated the RGB into three layers then stacked the input on the three layers before converting the output to grayscale as it produced outstanding results. Secondly, fitting large images of medical dataset was a big problem. The best approach was to utilize a high-end machine for training with at least 32GB Memory in RAM. Thirdly, deciding the layers in the VGG network to sample from for the loss function due to the difference in the dataset and choosing early layers will be considering more low features in the loss function and arrived at a sampling at the 9th layer as it presented the best loss function. Lastly, using trial and error we were able to determine the best patch size in the discriminator and arrived at a patch size of 16 as it produced the best performance on all dataset.

7.2 SRCNN

In the SRCNN model, there are only three parts: patch extraction and representation, non-linear mapping, and reconstruction. The patch extraction and representation involve upscaling the low-resolution input to the desired size using bicubic interpolation. After that, a non-linear mapping is performed to map a low-resolution vector to a high-resolution vector. Finally, the image is reconstructed in the last phase.

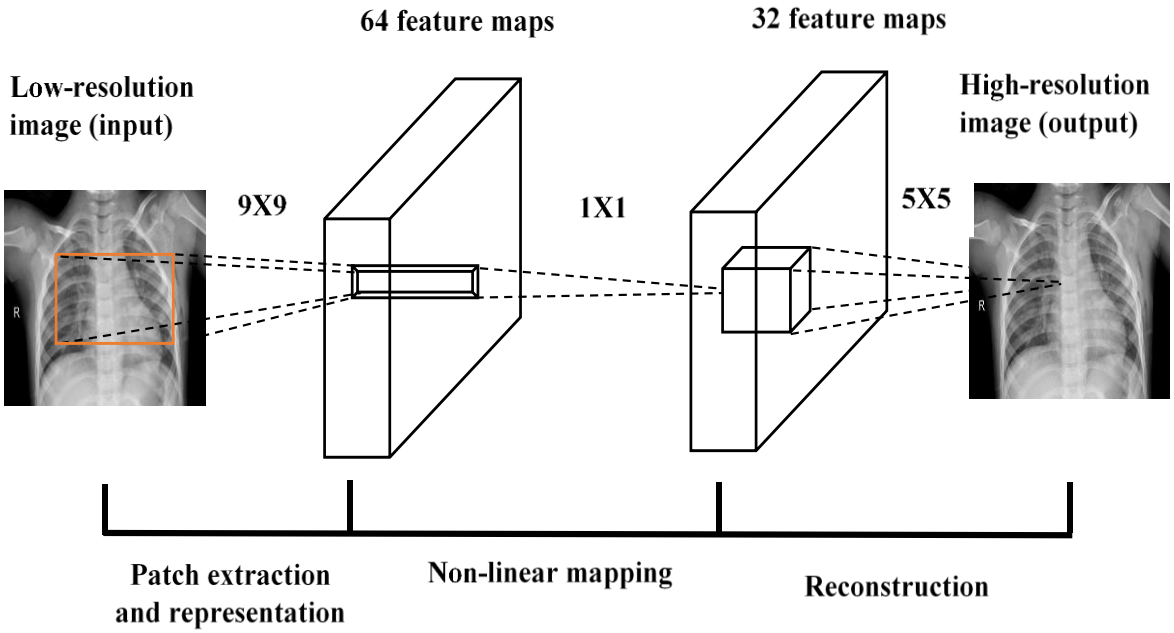


Figure 50. The filters map from low resolution to high resolution

As shown in Figure 50, we also applied some modifications to the SRCNN Architecture in order to work with the medical dataset. This involved changing the RGB images to YCrCb then using the first channel only as most of the visual data needed was on the first channel. Further updated the 32 filters of a size of 1*1 to 3*3, using 1 (in all n-channels) filter of size 5*5 to produce final HR patch while excluding padding as we planned the size to decrease as we target HR images from LR ones. We incorporated glorot uniform initializer and trained on patches of small sizes (32*32 or 64*64) by breaking up the images into patches but evaluating on the whole image as it made the network work on high-level Features.

7.3 Data Collection

Three datasets were employed in this research; CelebA, Pneumonia, and NIH chest x-ray images. The CelebA included about 202,599 face images in batches of 20 images from approximately 10,000 celebrities; it provided a set of pre-cropped face images aligned using the hand-labeled key points [70]. The Pneumonia chest x-ray set consisted of 5,863 images grouped into two classes, Pneumonia, and Normal patients. The images were extracted from the pediatric patient as part of patients' routine clinical care ranging from one year to five years at Guangzhou Women and Children's Medical Center in Guangzhou, China [59]. Lastly, the NIH chest x-ray dataset contains 112,120 chest x-ray images having disease labels from 30,805 unique patients. There are 15 classes: Atelectasis, Fibrosis, Effusion, Infiltration, Pneumothorax, Edema, Pneumonia, Pleural-thickening, Cardiomegaly, Nodule Mass, Consolidation, Emphysema, Hernia and one for No findings [54].

7.4 Experimental Setup

We coded all models in Python 3.6.7, Conda 4.5.11, Tensorflow 1.12.0, and Keras 2.1.6. The experiment was run on Intel® Core™ i7-3770 CPU @ 3.60GHz × 8, 64GB RAM, 64-bit, and Nvidia® TITAN XP 12GB. For the evaluation, we divided the data into 70% for training and 30% for testing. Each model was trained for 30,000 epochs, and at every 500 epochs, the validation set was evaluated on 20% of random samples removed from the training set. The test parameters used was an Adam optimizer, beta1 (momentum of Adam optimizer) = 0.5, learning rate = 0.0003. Mini-batch sizes range from 1, 32, to 128 (i.e., each epoch contains only one batch).

7.5 Result and Evaluation

The quantitative metrics employed in this study are PSNR, SSIM, and MSE. The SSIM as shown in Figure 51 is based on the computation of three terms, namely the luminance, contrast, and structural term. The overall SSIM index is a multiplicative combination of all three indices. It is independent of each image alone so as we can see that the extended SRCNN averaged around 0.7 for CelebA, 0.66 for the Pneumonia and 0.6 for NIH presenting the better result in comparison to both the CelebA and Pneumonia. For the extended SRGAN Analysis, the SSIM average around 0.3 for CelebA, 0.25 for the Pneuomnia and 0.6 for NIH. Which is very high compared to CelebA

and Pneumonia proving that the generator generated data with high accuracy showing that the discriminator was trained adequately.

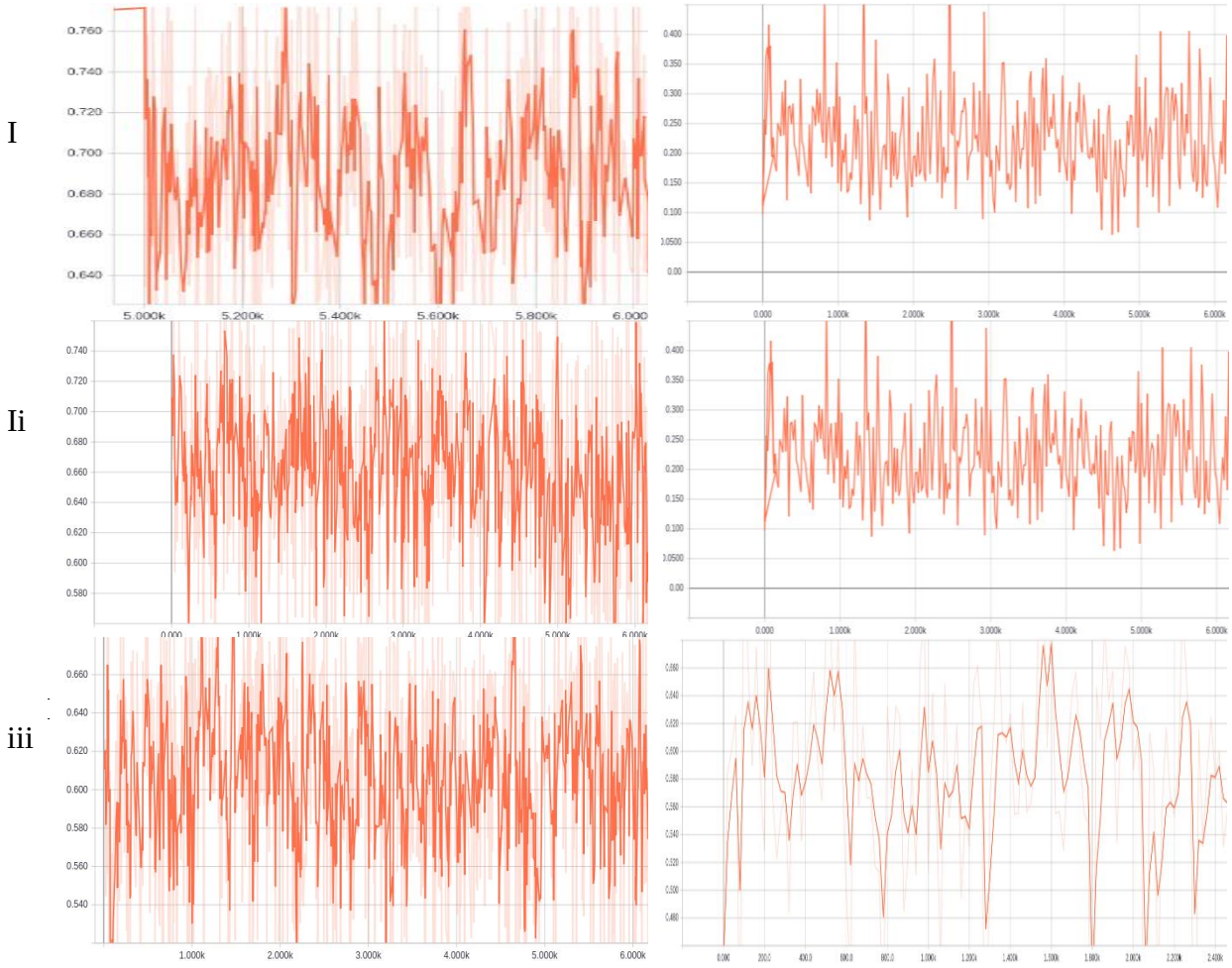


Figure 51. SSIM analysis of a) SRCNN, and b) SRGAN for all dataset

The PSNR is computed using the quantitative image restoration quality and calculated over a random sample of the dataset. For the CelebA set, the extended SRCNN result is around 28.0 while the srgan averaged at 9.0. While with the NIH chest x-ray dataset the extended SRCNN is about 38.0 but the extended SRGAN's output is around 13.0 achieving a better result than on the colored images in CelebA dataset with higher resolution. For the NIH data, the result of the extended SRCNN is around 39.0 higher than the 13.5 achieved by SRGAN including the colored images in CelebA dataset and chest x-ray dataset as shown in Figure 52.

As seen in Figure 53, for the CelebA set; the loss (MSE over patches) is decreasing indicating that the model is learning to produce HR images with a slight loss. The loss for extended SRCNN peaked around 0.0019, while the MSE averaged on 10.3 for extended SRGAN. Most of the

samples are low which means we have a good converge. For the NIH chest x-ray set, the loss reached about 0.0016 for the extended SRCNN. While the MSE peaked around 10.1 for the extended SRGAN which is less than the CelebA dataset meaning the technique performed better on the medical dataset, even though the resolution was higher. The loss reached about 0.0001 for the extended SRCNN with the MSE averaging around 10.1 for the extended SRGAN which is less than the CelebA dataset but similar to the chest x-ray result.

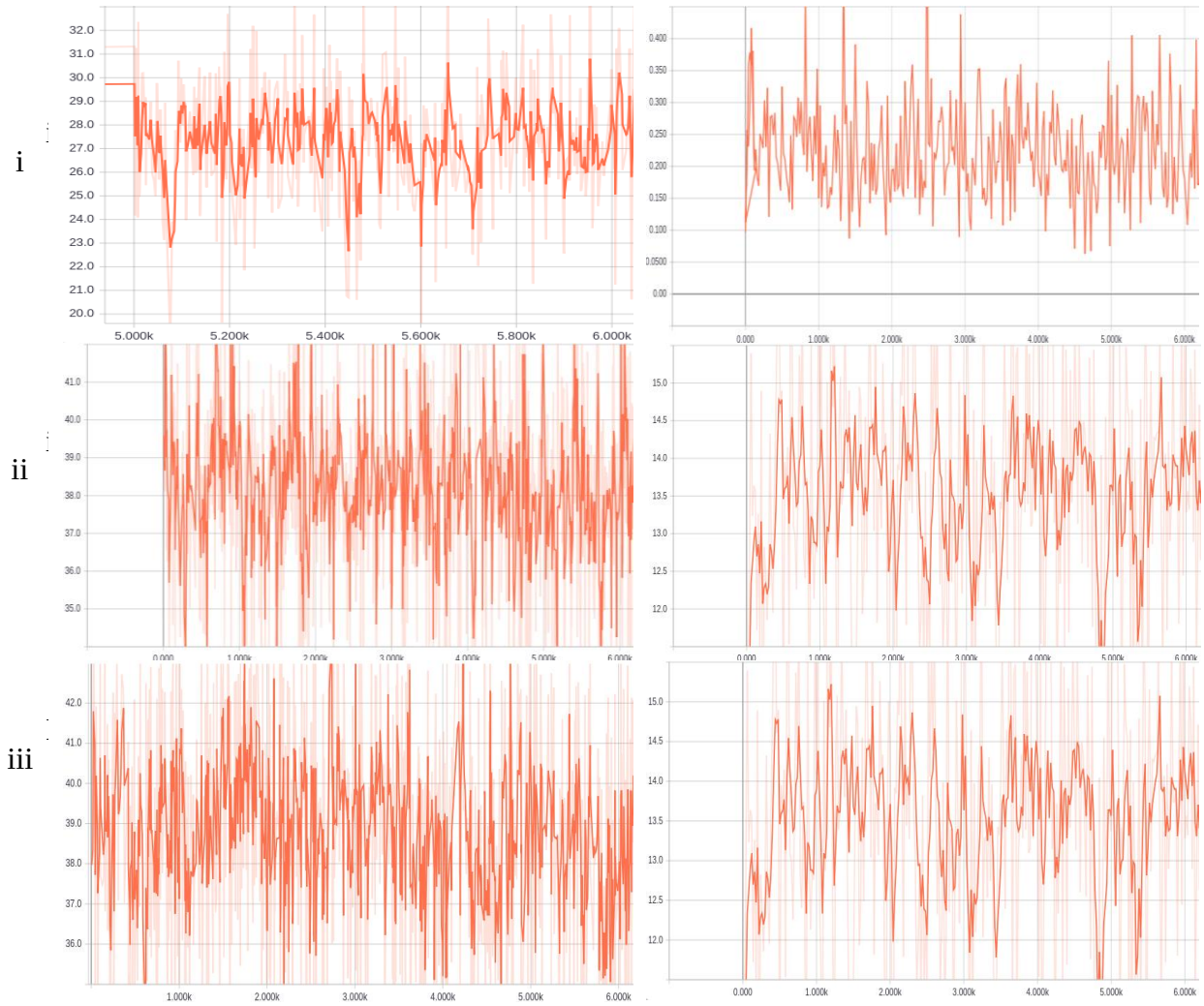


Figure 52. PSNR analysis of a) SRCNN, and b) SRGAN for all dataset

Table 24 shows the SSIM and PSNR on the different techniques on for CelebA, Pneumonia and NIH chest x-ray dataset. This SSIM and PSNR score was calculated on a random sample from the test set, and the SRCNN_GAN method was able to give a better SSIM score than the SRCNN and SRGAN methods while the PSNR of the SRCNN_GAN method was also higher than the SRGAN

and SRCNN methods for all dataset. In this case, quality was measured concerning edge fidelity. Interestingly, the models output perceptual accuracy is reflected in the PSNR and SSIM metrics. This was achievable the outputs from the SRCNN_GAN were visually closer to the original HR ground truth images. The SRGAN output images contain high-frequency information that is like the HR ground truth images while the SRCNN and SRGAN models tend to smooth and brighten the images out to achieve a high PSNR however due to the deblurring function it failed to outperform SRCNN_GAN. Lastly, the discriminator network seeks out the high-frequency information that differentiates HR and LR images, thus forcing the SRCNN_GAN output to have far more high-frequency details than the output of the SRGAN.

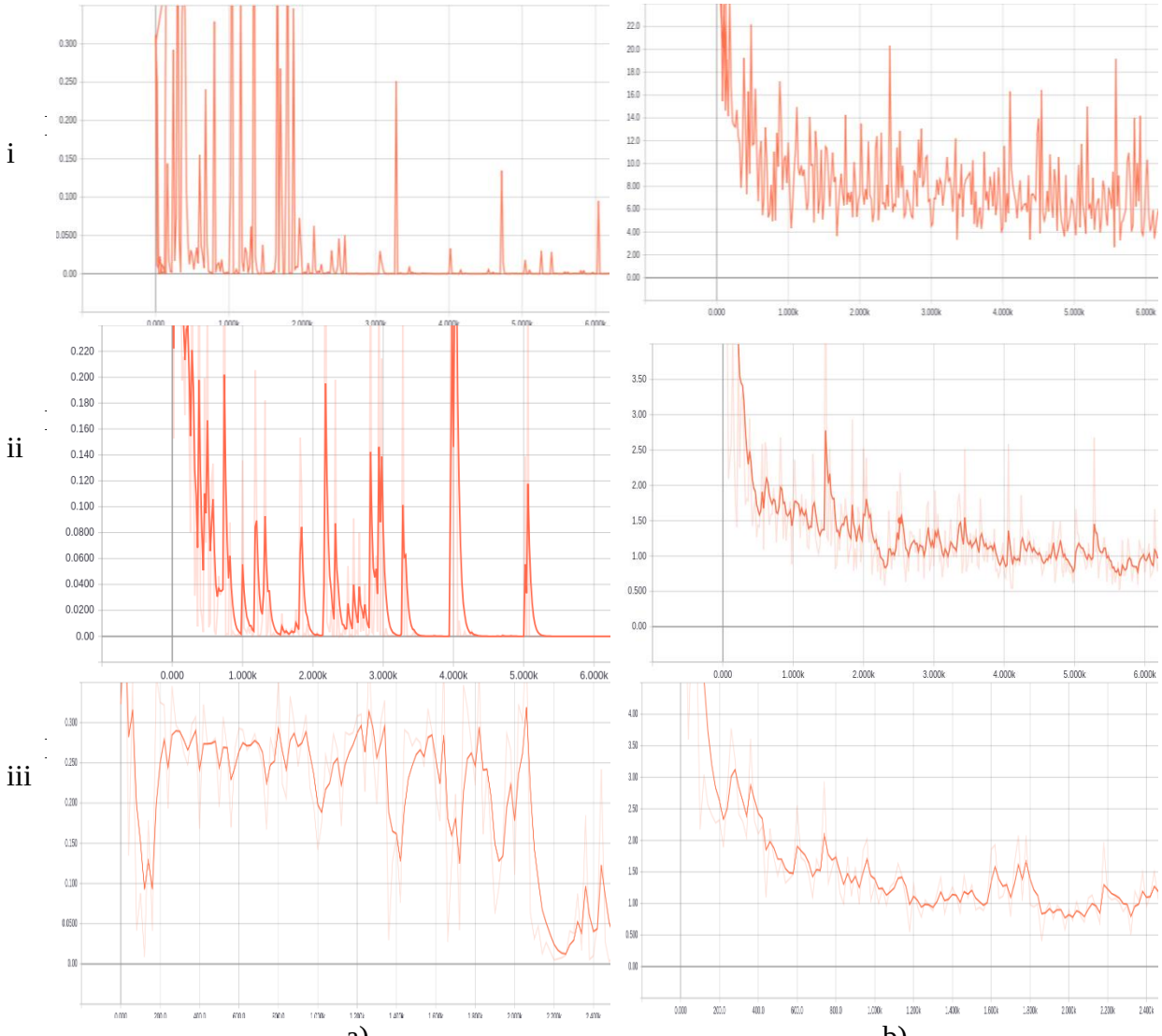


Figure 53. MSE loss evaluation of a) SRCNN, and b) SRGAN for all dataset

Table 24. Comparison of metrics on all models using all dataset

	CelebA		Pneumonia Chest x-ray		NIH Chest x-ray	
	SSIM	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**
SRCNN	0.545	14.51	0.890	32.81	0.832	26.18
SRGAN	0.579	14.49	0.888	35.83	0.895	31.26
SRGAN_CNN	0.917	25.87	0.992	41.93	0.991	38.36

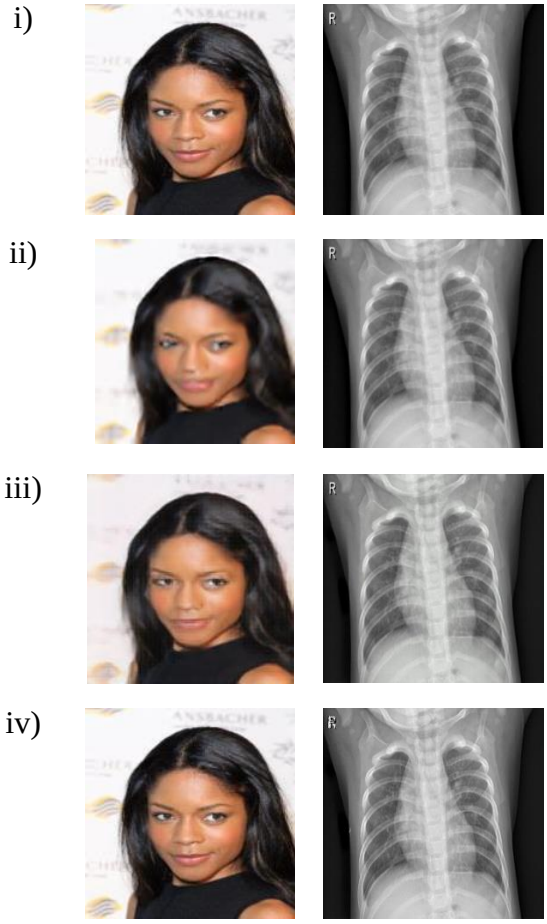


Figure 54. Sample i) ground truth, ii) SRCNN, iii) SRGAN, and iv) SRCNN_GAN model on dataset

Figure 54 shows an illustration of sample i) ground truth, ii) SRCNN, iii) SRGAN, and iv) SRCNN_GAN model trained on a) CelebA and b) NIH chest x-ray image. The SRCNN_GAN trains end-to-end on all dataset using a 64x64 input image, upscaled four times using the extended SRGAN to the 256x256 image before deblurring based on the extended SRCNN. Although the

SRCNN and SRGAN enhanced the image with high resolution, the change in brightness caused a problem in the dataset as the brightness preservation is necessary for the resolution optimization. However, in comparison with the ground truth, the SRCNN_GAN method results produced a sharper image.

Chapter 8

8 Conclusion

Banknote Identification is a relatively easy job for human eyes as the brain is capable of absorbing different kind of information and recognizing them with less effort. In image recognition, identification of banknote is a challenging task as a banknote can suffer from many defects once in circulation. Also. some images become skewed due to alignment and acquisition setup during the process of acquiring currencies via digital cameras or scanners. Much research has focused on banknote recognition using several feature extraction methods such as global features, local features, texture features, and edge-based features. Some papers have investigated both external (length, width, and thickness) and internal (color, texture) properties of the banknote. In the same view, global systems experimented with extracted features from the whole note such as Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF), Wavelet Transform (WT) and gray level properties. A banknote has also been classified using K-Nearest Neighbor (KNN), K-Means, Artificial Neural Network (ANN), Hidden Markov Model (HMM), and Support Vector Machine (SVM).

Smartphone usage has become part of our daily life, and its development is still snowballing. Most phones possess advanced functionalities that go beyond making calls and text messaging. They also serve as portable computing devices used in running e-mails, agendas, and storing data. Furthermore, mobile devices have at least one high-resolution camera that can be used to capture images and hold sufficient computational capability to process the image and execute algorithms for the recognition of banknotes. People identify currencies using several methods like inspecting the pictures, words, and numerals as well as distinctive signs, note consistency, and size. However, these techniques are prone to human error or omissions caused by visual impediments.

Physicians perform most interpretations of medical images; but, image analysis by humans is susceptible to error due to wide variations across interpreters, lethargy, and human subjectivity. Computer-aided diagnosis (CAD) is vital to improvements in medical analysis, as they facilitate

the identification of findings that need treatment and assist the expert's workflow. Several researchers have proposed incorporating CAD into the medical diagnostic process. It is a computerized procedure that offers a quantitative opinion to improve the clinical diagnostic process by providing an alternative view of medical image interpretation and diagnosis [119].

Over the last few years, deep learning systems have been introduced into medical image analysis domain with promising results on various applications, like a computerized diagnosis of breast cancer, mild cognitive impairment, blood cell classification, bone age estimation, Alzheimer's disease, radiology, organ segmentation and detection, and ultrasound standard plane selection. Thus, deep learning is rapidly proving to be the state-of-the-art system, leading to improved accuracy opening new frontiers in medical analysis with rates of progress not experienced before.

8.1 Summary of Contributions

In Chapter 3, we presented a method that extracted the Histogram of Gradient feature of the banknote, derived the principal components from the features and classified the samples using an MCSVM. A 10-fold cross-validation approach was employed in classifying the front and the backsides of the bill. Also, we extended the approach by proposing a Modified HOG/LVHCF descriptor for automatic recognition of paper money. It extracted the HOG descriptor from image patches created from the vertices of SURF interest points feature of the banknote, Low Variance, and High Correlation feature reduction filter was employed to reduce the feature set. It then classified the features using OVO-SVM in identifying and authenticating fake and genuine bills. Lastly, point matching technique was applied to detect feature vectors to match images containing multiple banknotes to determine the value of the currencies.

In Chapter 4, we described the development, implementation, and appraisal of a mobile phone banknote recognition system based on Unity 3D. The designed system is very general and with little modifications, it would work on any currency worldwide. The system uses several KNN classifier as base learners to sort data based on a calculated distance measure and employs Ensemble classifiers to combine the base KNN approach to create a strong classifier. It uses a two-tier prediction approach; the first level identifies the denomination of the banknote and the second level determines the authenticity of the currency. The system could recognize note values,

detect fake notes and reject banknotes that are not Naira notes. The proposed cascaded approach presented significantly higher recognition and detection accuracy than the base KNN classifiers with statistically significant improvement. Thus, it can be concluded that the proposed Cascaded system is superior to individual KNN, stacking and WMA method for banknote recognition and detection. Furthermore, it is essential to note that our proposed banknote recognition system only made use of features extracted from the Face value. In the same view, we studied the extracted features to determine its usefulness in determining the authenticity of a banknote after classifying the denomination of the bills. It further evaluated the robustness of the features by taking into consideration image variation like rotation, scaling, lighting and partial occlusions. The mobile-based system included an image acquisition and processing module built using Unity 3D, in addition to a voice output stating the denomination and confirming the authenticity of the Naira note. Additionally, the Backend MATLAB Server had the SIFT, SURF, and COMBINED descriptor extraction and a proposed classifier algorithm based on a stacked MLP-ECOCSVM classification and an authentication module. The system uses a three-tier approach; the first level identifies the denomination of the banknote, the second level determines the authenticity of the currency while the last stage rejects paper currencies that are not Naira, USD, and CAD notes. The proposed method employed a 10-fold Cross Validation and was tested on a randomized database sample comprising of counterfeit and real Naira note in addition to the US and Canadian dollars. The proposed system achieved a recognition accuracy of 97.18% and a detection rate of 95.45%. Moreover, we designed a hybrid system using Unity 3D as a front-end platform for a mobile phone application which includes a preprocessing module and backend server programmed using MATLAB for feature extraction, classification, and authentication system based on the HOG-HCF feature classified using WMA (MLPNN+MCSVM) has been designed. The proposed approach is robust to illumination changes and distortions as it was able to recognize partial, occluded and distorted banknotes irrespective of their orientation. The presented approach could identify all the denominations of the Naira, USD, CAD, and Euro notes. The system extracted the Histogram of Gradient feature, four feature reduction techniques were applied, and five classifiers employed to determine the denomination and authenticity of the banknote.

In Chapter 5, inspired by recent successes with deep learning techniques, we attempted to address the longstanding fundamental feature extraction problem using CNN ensemble for classification of medical images without computing the morphology and texture features. We

proposed a deep CNN ensemble to classify White blood cell and chest x-ray images. The network consists of 3 convolutional layers including Relu activations, followed by two max pooling in the 2nd and 3rd layers with one fully connected layer to output the learned features. The ensemble classifier employed five base classifiers aggregated using three ensemble learners with bagging presenting the best accuracy. The system architecture comprised Unity 3D as a front-end platform for the mobile phone application, and backend server was programmed using MATLAB for formatting, feature extraction, and classification of the medical image. The encouraging results of our research indicate that the proposed deep learning framework outperforms conventional hand-crafted feature computing CAD frameworks. As a result, this research serves as a basis to address the medical image diagnosis problem of both grayscale and colored images. It is believed that this is just the beginning for deep learning application to specific image recognition tasks and there are still several unattended challenges in the analysis of pathological data.

In Chapter 6, we implemented a BIGAN architecture to mitigate the effect of the BCE loss in a Semi-supervised setting. It employed a DCGAN autoencoder, a generative and encoding autoencoder in which the generative path is a Deep convolutional GAN that replaces all pooling layers with fractional-strided convolutions, includes batch normalization, Relu activation for all layers except the output layer that uses Tanh and removes fully connected hidden layers for the deeper architectures. While the discriminator explores an alternative to the conventional CNN approach called Capsule Network. It replaces the max-pooling with a routing dynamic routing mechanism and the scalar-output feature detectors employed by CNN with a vector-output Capsule but still utilizes translated replication of the trained feature detectors of the CNN as it allows translated knowledge of good weight values extracted at one position to other positions in the image. To achieve the reproducibility of learned knowledge across space, the Capsule network makes use of a convolutional layer in all the last layers. One of the main benefits of the Capsule network is the ability to deal with translation. It reduces exponential issues by converting pixel intensities into vectors of instantiation parameters of predictable fragments before applying transformation matrices to the regions to forecast the instantiation parameters of larger regions. It employs neural activities which have varying viewpoints thereby providing advantages over normalization methods like spatial transformer networks. Capsules also handle segmentation issues using the vector's routing-by-agreement of the vector of the instantiation parameter. The first experiment was conducted using CelebA and Pneumonia dataset; further experimental

evaluation was performed using NIH chest x-ray dataset. The system achieved significant improvements in generating photo-realistic images using a semi-supervised method that mitigates the common problems of global consistency, random dimension and translation, mode collapse, and stabilizing training of GANs.

In Chapter 7, deep learning based super-resolution upscaling, deblurring and reconstruction method for single image super-resolution was proposed. It involved redesigning the SRGAN, using VGG19 network for feature extraction, setting discriminator network's working space as feature space, and adding the loss function based on the mean square error of pixel space. The technique gains more details by incorporating SRCNN layers to increase the PSNR in the reconstruction at the same time. We demonstrated the efficacy of our method in the application of the CelebA dataset while further analysis was conducted for medical Chest Radiography enhancement. We employed visual assessment for the qualitative evaluation which confirmed that the extended Convolution based SRGAN generator produced much sharper edge than traditional SRGAN, SRCNN, and even the ground truth images. While the quantitative evaluation utilized three image quality metrics: SSIM, MSE, and PSNR. The results showed that the extended SRGAN scheme provided MSE, PSNR, and SSIM that were significantly better than the base SRGAN and SRCNN techniques. Lastly, several tradeoffs were explored using different network structures, parameter settings and finetuning to achieve the best performance and speed.

8.2 Future Directions

In the field of Banknote Recognition and Authentication, future work would include enlarging the size of the dataset to compare similar feature extraction techniques such as SURF and SIFT, with the proposed HOG based system and introducing bias-reducing features to the system to detect and reject counterfeit notes. In addition to incorporating new feature descriptors and reduction approach to reduce processing time and scale down the application to work on mobile devices, reducing the computational complexity to enable embedding in micro-controllers and microchips. We will also be implementing a feature fusion system that combines other security features such as security thread, serial number, micro-letters, signature, and face on the banknote to improve the detection rate of the system. Similarly, usability testing would be conducted with the participation of individuals with visual impairment. Lastly, we will introduce feature reduction

and classification approaches to scale down the application to work solely by the processing capability of the mobile devices.

In the area of Computer Aided Diagnosis, we will investigate more advanced deep learning techniques and evaluate diverse dataset for more in-depth empirical studies. Future research would include Transfer learning from the large-scale annotated natural image dataset (ImageNet) to CAD problems for medical image diagnosis and extending the method to consider three-dimensional data for analyzing medical images. Furthermore, we will implement a PixelGAN autoencoder to condition a low-resolution image to generate higher resolution images. This involves scaling the autoregressive generative models to large scale images while finding efficient ways to incorporate latent variables in these architectures by training end-to-end. Furthermore, we will combine PixelGAN with Extreme Learning Machine (Classification) to solve the problem of conditioning a low-resolution image to generate higher resolution images and compare their performance with each other. Lastly, we would involve the other deep architectures like Stacked denoising sparse autoencoder (SDSAE) that can effectively eliminate the influence of noise. We will consider combining SDSAE with (Extreme Learning Machine) ELM to solve this problem and compare their performance with each other and consider the clustering task in the last step of our approach.

References

1. A. Agarwal, and S. Goswami, “An Efficient Algorithm for Automatic Car Plate Detection and Recognition,” In Proc. of the Second International Conference on Computational Intelligence & Communication Technology (CICT), pp. 644-648, February 2016, Ghaziabad, India.
2. H. Aggarwal, and P. Kumar, “Indian Currency Note Denomination Recognition in Color Images,” In International Journal on Advanced Computer Engineering and Communication Technology, Vol. 1, No. 1, pp. 12-18, 2014.
3. F. Ahangaryan, T. Mohammadpour, and A. Kianisarkaleh, “Persian Banknote Recognition Using Wavelet and Neural Network,” In Proc. of the International Conference on Computer Science and Electronics Engineering (ICCSEE), Vol. 3, pp. 679-684, 2012, Hangzhou, China.
4. A. Ahmadi, S. Omatu, T. Fujinaka, and T. Kosaka, “Improvement of reliability in banknote classification using reject option and local PCA,” In Information Sciences, Vol. 168, No. 1, pp. 277-293, 2004.
5. Z. Ahmed, S. Yasmin, M.N. Islam, and R.U. Ahmed “Image processing based Feature extraction of Bangladeshi banknote,” In Proc. of the 8th International Conference of Software, Knowledge, Information Management and Applications (SKIMA), pp. 1-8, 2014, Dhaka, Bangladesh.
6. J. Allebach, and P. W. Wong “Edge-directed interpolation,” In Proc. of the International Conference on Image Processing (ICIP), pp. 707–710, 1996, Lausanne, Switzerland.
7. S. Angsupanich, and S. Matayong, “Applying the mental model for real-time recognition of Thai banknote: The blinds' mobile application,” In Proc. of the 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 86-90, 2016, Kuala Lumpur, Malaysia.
8. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, “Lung pattern classification for interstitial lung diseases using a deep convolutional neural network,” In IEEE transactions on medical imaging, Vol. 35, No. 5, pp. 1207-1216, 2016.
9. P.O. Awe, and S. Lambotharan, “Cooperative spectrum sensing in cognitive radio networks using multi-class support vector machine algorithms,” In Proc. of the 9th

International Conference on Signal Processing and Communication Systems (ICSPCS), pp. 1-7, 2015, Cairns, QLD, Australia.

10. Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," In Proc. of the International Symposium on Biomedical Imaging (ISBI), pp. 294-297, April 2015, New York, USA.
11. C. Barillot, and N. Navab, "Medical Image Computing and Computer-Assisted Intervention," MICCAI 2013, Berlin, Germany, Springer, 2013, Nagoya, Japan.
12. H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool, "SURF: speeded up robust features," In Comput. Vis. Image Underst., Vol. 110, No. 3, pp. 346–359, 2008.
13. A. Berenguel, O.R. Terrades, J. Lladós, and C. Cañero, "Banknote counterfeit detection through background texture printing analysis," In Proc. of the 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 66-71, 2016, Santorini, Greece.
14. D. Berthelot, T. Schumm, and L. Metz, "BEGAN: boundary equilibrium generative adversarial networks," In Proc. of the International Conference on Machine Learning, Cornell University, 2017, arXiv preprint arXiv:1703.10717.
15. C. Brestel, R. Shadmi, I. Tamir, M. Cohen-Sfaty and E. Elnekave, "RadBot-CXR: Classification of Four Clinical Finding Categories in NIH chest x-ray dataset using Deep Learning," In Proc. of the Medical Imaging with Deep Learning Conference, 2018, Amsterdam, Netherland.
16. J. Bruna, P. Sprechmann and Y. LeCun, "Super-resolution with deep convolutional sufficient statistics," In Proc. of the International Conference on Learning Representations (ICLR), 2016, San Juan, Puerto Rico.
17. W. Chen, S. Ding, Z. Chai, D. He, W. Zhang, G. Zhang, and W. Luo, "FPGA-Based Parallel Implementation of SURF Algorithm," In Proc. of the 22nd International Conference on Parallel and Distributed Systems (ICPADS), pp. 308-315, 2016, Wuhan, Hubei, China.
18. Y. Chen, Z. Lai, J. Wen and C. Gao, "Nuclear norm based two-dimensional sparse principal component analysis," In International Journal of Wavelets, Multiresolution and Information Processing, Vol. 16, No. 2, p. 1840002, 2018.
19. S. Chen, L. Liang, W. Liang, and H. Foroosh, "3d pose tracking with multi template warping and sift correspondences," In IEEE Transactions on Circuits and Systems for

Video Technology, Vol. 26, No. 11, pp. 2043-2055, 2016.

20. B.V. Chetan, and P.A. Vijaya, "A Robust Side Invariant Technique of Indian Paper Currency Recognition," In International Journal of Engineering Research and Technology, Vol. 1, No. 3, pp. 1-7, ESRSA Publications, 2012.
21. E. Choi, J. Lee, and J. Yoon, "Feature extraction for banknote classification using wavelet transform," In Proc. of the 18th International Conference on Pattern Recognition (ICPR), pp. 934-937, 2006, Hong Kong, China.
22. Chollet, F. "Keras: Deep learning library for theano and tensorflow," URL: https://keras. io/k,7(8).
23. K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, "An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm," In Proc. of the 3rd International Conference on Industrial Application Engineering 2015 (ICIAE2015), 2015, Kitakyushu, Japan.
24. CNIB, Canadian Bank Note Reader. Retrieved from http://www.cnib.ca/en/services/products/bank-note-reader, 1996.
25. R. Dahl, M. Norouzi, and J. Shlens, J. "Pixel recursive super resolution," In Proc. of the International Conference on Computer Vision (ICCV), Vol. 1, No. 2, 2017, Venice, Italy.
26. B.P. Daniel and C.U. Idoko "The Macro-Economic Consequences and Regulatory Challenges of Currency Counterfeiting In Nigeria," In International Journal of Financial Economics Vol. 3, No. 2, pp. 113-120, 2014.
27. L. Deng, "The CelebA database of handwritten digit images for machine learning research," In IEEE Signal Processing Magazine, Vol. 29, No. 6, pp. 141-142, 2012.
28. G. Dişken, Z. Tüfekçi, L. Saribulut and U. Çevik, "A review on feature extraction for speaker recognition under degraded conditions," In IETE Technical Review, Vol. 34, No. 3, pp. 321-332, 2017.
29. J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," Machine Learning, Cornell University, 2016, arXiv preprint arXiv:1605.09782.
30. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38, No. 2, pp. 295–307, 2016.

31. Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V.C. Mok, L. Shi, and P.A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," In IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1182–1195, May 2016.
32. I.A. Doush, and A.B. Sahar, "Currency recognition using a smartphone: Comparison between color SIFT and grayscale SIFT algorithms," In Journal of King Saud University-Computer and Information Sciences, 2016.
33. A. Dosovitskiy, and T. Brox, "Generating images with perceptual similarity metrics based on deep networks" In Advances in Neural Information Processing Systems, pp. 658-666, 2016.
34. U. K. Durgam, S. Paul, and U. C. Pati, "SURF based matching for SAR image registration," In Proc. of the Students Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-5, March 2016, Bhopal India.
35. J. G. Elmore, G. M. Longton, P. A. Carney, B. M. Geller, T. Onega, A. N. Tosteson, H. D. Nelson, M. S. Pepe, K. H. Allison, S. J. Schnitt, and F.P. O'Malley, "Diagnostic concordance among pathologists interpreting breast biopsy specimens," In Journal of the American Medical Association (Jama), Vol. 313, No. 11, pp. 1122–1132, 2015.
36. S. Gai, G. Yang, and M. Wan, "Employing quaternion wavelet transform for banknote classification," In Neurocomputing, Vol. 118, pp. 171-178, 2013.
37. F. García-Lamont, J. Cervantes, and A. López, "Recognition of Mexican banknote via their color and texture features," In Expert Systems with Applications, Vol. 39, No. 10, pp. 9651-9660, 2012.
38. F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," In Annual Review of Pathology: Mechanisms of Disease, vol. 8, pp. 331–359, 2013.
39. D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," In Proc. Of the 12th International Conference in Computer Vision, pp. 349-356, 2009, Kyoto, Japan.
40. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ... & Y. Bengio, "Generative adversarial nets," In Advances in neural information processing systems," pp. 2672-2680, 2014.

41. A. Govada, G. Bhavul, K.S. Sanjay, “Distributed multi-class SVM for large dataset,” In Proc. of the Third International Symposium on Women in Computing and Informatics, pp. 54-58, 2015, Kochi, India.
42. M. Govindarajan "Evaluation of ensemble classifiers for intrusion detection," World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol. 10, No. 6, pp. 876-884, 2016.
43. S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, “Convolutional sparse coding for image super-resolution,” In Proc. of the International Conference on Computer Vision (ICCV), pp. 1823– 1831, 2015, Las Condes, Chile.
44. D. A. K. S. Gunaratna, N. D. Kodikara, and H. L. Premaratne. "ANN based currency recognition system using compressed grayscale and application for Sri Lankan currency note-SLCRec," In International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol. 2, No. 9, pp. 235-240, 2008.
45. J. Guo, Y. Zhao, and A. Cai, “A reliable method for paper currency recognition based on LBP,” In Proc. of the International Conference on Network Infrastructure and Digital Content, pp. 359-363, 2010, Beijing, China.
46. M. Habibzadeh, A. Krzyżak, and T. Fevens, “Comparative study of shape, intensity and texture features and support vector machine for White blood cell classification,” In Journal of Theoretical and Applied Computer Science, Vol. 7, No. 1, pp. 20-35, 2013.
47. M. Hasanuzzaman, X. Yang, and Y. Tian “Robust and effective component-based banknote recognition for the blind. In IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 42, No. 6, 2012, pp. 1021-1030.
48. H. Hassanpour, A. Yaseri, and G. Ardeshiri, “Feature extraction for paper currency recognition,” In Proc. Of the 9th International Symposium on Signal Processing and Its Applications (ISSPA), pp. 1-4, 2007, Sharjah, United Arab Emirates.
49. N. Hatami, and S. Seyedtabaii, “Error correcting output codes using genetic algorithm-based decoding,” In Proc. of the Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on IEEE, Vol. 1, pp. 391-396, 2008, Gyeongju, South Korea.
50. J. He, H. Zhang, O. Jin, Z. Hu, and J. Zhang, J. “A Novel Recognition Method for Nigeria

Paper Currency based on Support Vector Machine & Genetic Algorithms," International Symposium on Signal Processing, Biomedical Engineering and Informatics (SPBEI 2013), Vol. 6, pp. 1087-1094, 2013, Hangzhou, China.

51. C. Herley, P. Vora, and S. Yang, Detection and deterrence of counterfeiting of valuable documents, In Proc. of the International Conference on Image Processing (ICIP'04), pp. 2423-2426, 2004, Singapore, Singapore.
52. J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5197–5206, 2015, Boston, Massachusetts, USA.
53. T. Ishigaki and T. Higuchi, Detection of Worn-out Banknote by Using Acoustic Signals, Transactions of the Society of Instrument and Control Engineers, Vol. 44, No. 5, pp. 444-449, 2008.
54. S. Jaeger, S. Candemir, S. Antani, Y. X. J. Wáng, P. X. Lu, and G. Thoma, "Two public chest x-ray dataset for computer-aided screening of pulmonary diseases," In Quantitative imaging in medicine and surgery, Vol. 4, No. 6, pp. 475, 2014.
55. J. Johnson, A. Alahi and F. Li, "Perceptual losses for real-time style transfer and super-resolution," In Proc. of the European Conference on Computer Vision (ECCV), pp. 694–711. Springer, 2016, Amsterdam, Netherland.
56. H. Jung, Y. Ehara, J. K. Tan, H. Kim, and S. Ishikawa, "Applying MSC-HOG Feature to the Detection of a Human on a Bicycle," In Proc. of the 12th International Conference Control, Automation and Systems (ICCAS), pp. 514-517, 2012, JeJu Island, South Korea.
57. R. Kadota, H. Sugano, M. Hiromoto, H. Ochi, R. Miyamoto, and Y. Nakamura, "Hardware architecture for HOG feature extraction," In Proc. of the Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'09), pp. 1330-1333, 2009, Kyoto, Japan.
58. N. Kato, and S. Omachi, "A handwriting character recognition system using directional element feature," In IEEE Trans. Pattern Anal. Mach. Intell, Vol. 21, No. 3, pp. 258–262, 1999.
59. D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and NIH chest x-ray datasetImages for Classification," Mendeley Data, v2, 2018, Available: http://dx.doi.org/10.17632/rscbjbr9sj.2.

60. J. Kim, L. Kwon, J. Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks” In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 1646-1654, 2016, Nevada, USA.
61. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” In Advances in neural information processing systems, pp. 1097-1105, 2012.
62. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, and A. Acosta, A.P. Aitken, A. Tejani, J. Totz, Z. Wang and W. Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” In Proc. of the IEEE conference on computer vision and pattern recognition (CVPR), Vol. 2, No. 3, p. 4, 2017, Hawai, USA.
63. S. Lee, H. Son, J. C. Choi, and K. Min, “HOG feature extractor circuit for real-time human and vehicle detection,” In Proc. of the TENCON 2012-2012 IEEE Region 10 Conference, pp. 1-5, 2012, Cebu, Philippines.
64. X. Li and M. Orchard “New edge-directed interpolation,” In IEEE Transactions on Image Processing (TIP), Vol. 10, pp. 1521–1527, 2001.
65. J. Li, Z. Zhan, C. Ding, and Y. Wang, “FaceGANs: Stable Generative Adversarial Networks with High-Quality Images,” In Proc. of the Sixth International Conference on Learning Representations, 2018, Vancouver, Canada.
66. Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L. J. Li, and F.F. Li, “Thoracic disease identification and localization with limited supervision,” In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, Utah, USA.
67. Y. Liang, L. Liu, Y. Xu, Y. Xiang, and B. Zou, “Multi-task gloh feature selection for human age estimation,” (ICIP), In Proc. of the 18th IEEE International Conference on Image Processing, pp. 565-568, 2011, Brussels, Belgium.
68. B. Lin, and B. Fang, “A New Spatial-temporal Histograms of Gradients Descriptor and HOD-VLAD Encoding for Human Action Recognition,” International Journal of Wavelets, Multiresolution and Information Processing, 2018.
69. X. Liu, “A camera phone-based currency reader for the visually impaired,” In Proc. of the 10th international ACM SIGACCESS conference on Computers and accessibility ACM, pp. 305-306, 2008, Halifax, Canada.

70. Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” In Proc. of the International Conference on Computer Vision, pp. 3730–3738, 2015, Las Condes, Chile.
71. J. F. Liu, S. B. Liu, and X. L. Tang, “Long, An algorithm of real-time paper currency recognition,” In J. Comput. Res. Dev. Vol. 40, No. 7, pp. 1057–1061, 2003.
72. J. Lee, J. Seong-Goo, K. II-Hwan, “Distinctive point extraction and recognition algorithm for various kinds of euro banknote,” In International Journal of Control, Automation, and Systems, Vol. 2, No. 2, pp. 201-206, 2004.
73. A. Makhzani and B. J. Frey, “Pixelgan autoencoders,” In Advances in Neural Information Processing Systems, pp. 1972–1982, 2017.
74. X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, and S.P. Smolley, “Least squares generative adversarial networks,” In Proc. of the IEEE International Conference on In Computer Vision (ICCV), pp. 2813-2821, 2017, Venice, Italy.
75. M. Matsumoto, “Parameter setting of self-quotient ε-filter using HOG feature distance,” In Proc. of the Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), pp. 127-133, April 2011, Paris, France.
76. L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, “Unrolled generative adversarial networks,” Machine Learning, Cornell University, 2016, arXiv preprint arXiv:1611.02163.
77. R. Mikut, O. Burmeister, S. Braun, and M. Reischl, "The open-source Matlab toolbox Gait-CAD and its application to bioelectric signal processing," In Proc. of the DGBMT-Workshop Biosignal, pp. 109-111, 2008, verarbeitung, Potsdam, Germany.
78. T. Mooney, “Kaggle dataset for single WBC blood cell images,” 2018, Available: https://www.kaggle.com/paultimothymooney/blood-cells.
79. S.E.A. MrigankaGogoi, and S. Mukherjee, “Automatic Indian Currency Denomination Recognition System based on Artificial Neural Network,” In Proc. of the International Conference on Signal Processing and Integrated Networks (SPIN), pp. 553 – 558, 2015, Noida, India.
80. D. Mulmule-Shirkhedkar, and A. R. Dani, “Comparative study of SURF and FREAK descriptor on Indian Rupee Currency note,” In Proc. of the International Conference on Information Processing (ICIP), pp. 784-789, December 2015, Pune, India.
81. A.M. Murugavel, S. Ramakrishnan, U. Maheswari, and B.S. Sabetha, “Combined seizure

index with adaptive multi-class SVM for epileptic EEG classification," In Proc. of the International Conference on Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT), pp. 1-5, January 2013, Tiruvannamalai, India.

82. S. Nanayakkara, R. Shilkrot, and P. Maes, "EyeRing: A finger-worn assistant," In CHI'12 extended abstracts on human factors in computing systems, pp. 1961–1966, May 2012, Vienna, Austria.
83. M. Olgun, A. O. Onarcan, K. Özkan, S. Işik, O. Sezer, K. Özgişi, and O. Koyuncu, "Wheat grain classification by using dense SIFT features with SVM classifier," In Computers and Electronics in Agriculture, Vol. 122, pp. 185-190, 2016.
84. Orbit Research, iBill, Retrieved from http://www.orbitresearch.com/ibill_details.php, 2013.
85. N.S. Pai, J.H. Hong, P.Y. Chen, and J.K. Wu, "Application of design of image tracking by combining SURF and TLD and SVM-based posture recognition system in robbery pre-alert system," In Multimedia Tools and Applications, Vol. 76, No. 23, pp. 25321-25342, 2017.
86. N. Paisios, A. Rubinsteyn, V. Vyas, L. Subramanian, L. "Recognizing currency bills using a mobile phone: an assistive aid for the visually impaired," In Proc. of the 24th annual ACM symposium adjunct on User interface software and technology ACM, pp. 19-20, 2011, New York, USA.
87. P.D. Pawar, and S.B. Kale, "Recognition of Indian Currency Note Based on HSV Parameters," In International Journal of Science and Research (IJSR), Vo. 3, No.6, pp. 132-137, 2014.
88. T.D. Pham, Y.H. Park, S.Y. Kwon, K.R. Park, D.S. Jeong, and S. Yoon, "Efficient Banknote Recognition Based on Selection of Discriminative Regions with One-Dimensional Visible-Light Line Sensor," In Sensors, Vol. 16, No. 3, pp. 328, 2016.
89. R. Polikar, "Ensemble learning," In Ensemble machine learning, pp. 1-34, Springer, Boston, MA, 2012.
90. J. Prinyakupt and C. Pluempitiwiriyawej "Segmentation of White blood cells and comparison of cell morphology by linear and naïve Bayes classifiers," In Biomedical engineering online, Vol. 14, No. 1, p. 63, 2015.

91. A. Radford, L. Metz, and S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, Machine Learning, Cornell University, 2015, arXiv preprint arXiv:1511.06434.
92. S. Rahman, P. Banik, and S. Naha, S. "LDA based paper currency recognition system using edge histogram descriptor," In Computer and Information Technology (ICCIT), In 17th International Conference on IEEE, pp. 326-331, 2014.
93. H. Rajput, T. Som, and S. Kar, "Vehicular License Plate Localization Using Principal Component Analysis," In IETE Technical Review, Vol. 34, No. 2, pp. 214-219, 2017.
94. A. Rashid, A. Prati, and R. Cucchiara, "On the design of embedded solutions to banknote recognition," In Optical Engineering, Vol. 52, No. 9, 2013.
95. R. Rathee, "A Review Paper on Currency Recognition System," In Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. Org, Vol. 6, No. 8, 2016.
96. T. Reiff, and P. Sincak, "Multi-Agent Sophisticated System for Intelligent Technologies," In Proc. of IEEE 6th International Conference on Computational Cybernetics, 2008, Stara Lesna, Slovakia.
97. H. Ren, M. El-Khamy and J. Lee, "CT-SRCNN: cascade trained and trimmed deep convolutional neural networks for image super-resolution," In Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1423-1431, 2018, Lake Tahoe, USA.
98. Y. Ren, M. Nguyen, and W.Q. Yan, "Real-Time Recognition of Series Seven New Zealand Banknote," In International Journal of Digital Crime and Forensics (IJDCF), Vol. 10, No. 3, pp. 50-65, 2018.
99. H.R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R.M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," In IEEE Trans. Med. Imag., Vol. 35, No. 5, pp. 1170–1181, May 2016.
100. A. Roy, B. Halder, and U. Garain, "Authentication of currency note through printing technique verification," In Proc. of the 7th Indian Conference on Computer Vision, Graphics and Image Processing, pp. 383-390, 2010, Chennai, Dubai.
101. S. Sabour, N. Frosst, and G.E. Hinton, "Dynamic routing between Capsules," In Advances in Neural Information Processing Systems, pp. 3856-3866, 2017.

102. V.D. Sachdeva, E. Fida, J. Baber, M. Bakhtyar, I. Dad, and M. Atif, "Better object recognition using bag of visual word model with compact vocabulary," In Proc. of the 13th International Conference on Emerging Technologies (ICET), pp. 1-4, 2017, Islamabad, Pakistan.

103. B. Sahiner, H.P. Chan, M.A. Roubidoux, L.M. Hadjiiski, M.A Helvie, C. Paramagul, J. Bailey, A.V. Nees, and C. Blane, "Malignant and Benign Breast Masses on 3D US Volumetric Images: Effect of Computer-aided Diagnosis on Radiologist Accuracy 1," In Radiology, Vol. 242, pp. 716–724, 2007.

104. H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks," In Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, Calgary, Canada.

105. K. Sathisha, "Bank automation system for Indian currency-a novel approach," In 2011 IEEE Recent Advances in Intelligent Computational Systems, 2011.

106. J. Schmidhuber, "Deep learning in neural networks: An overview," In Neural networks, Vol. 61, pp. 85-117, 2015.

107. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. Van Riel, M. Wille, M. Naqibullah, C. Sanchez and B. Van Ginneken, "Pulmonary nodule detection in CT images using Multiview convolutional networks," In IEEE Trans. Med. Imag., Vol. 35, No. 5, pp. 1160–1169, May 2016.

108. S. Singh, S. Choudhury, K. Vishal, and C.V. Jawahar, "Currency Recognition on Mobile Phones," In Proc. of the 22nd International Conference Pattern Recognition (ICPR), pp. 2661-2666, 2014, Stockholm, Sweden.

109. K. Sirinukunwattana, S.E.A. Raza, Y.W. Tsang, D. R. Snead, I.A. Cree, and N.M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," In IEEE Trans. Med. Imag., Vol. 35, No. 5, pp. 1196–1206, May 2016.

110. C. Stein, C. Nickel, and C. Busch, "Finger-photo recognition with smartphone cameras," In Proc. of the International Conference on Biometrics Special Interest Group (BIOSIG), pp. 1-12, 2012, Darmstadt, Germany.

111. M.C. Su, C.Y. Cheng, and P.C. Wang, “A neural-network-based approach to White blood cell classification,” In The scientific world journal, 2014.

112. H. I. Suk. and D. Shen, “Deep learning-based feature representation for AD/MCI classification,” In Proc. of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 583-590, Springer, September 2013, Berlin, Heidelberg.

113. Y. Sun, X. Wang and X. Tang, “Deep Learning Face Representation by Joint Identification-Verification,” In Advances in neural information processing systems, Technical report, 2014.

114. F. Takeda, L. Sakoobunthu and H. Satou, “Thai banknote recognition using neural network and continues learning by DSP unit,” In Knowledge-Based and Intelligent Information and Engineering Systems, pp. 1169-1177, 2003.

115. N. Theera-Umpon, and S. Dhompongsa, “Morphological granulometric features of nucleus in automatic bone marrow White blood cell classification.” In IEEE Transactions on Information Technology in Biomedicine, Vol. 11, No. 3, pp. 353-359, 2007.

116. R. Timofte, V. De Smet, and L. Van Gool, “Anchored neighborhood regression for fast example-based super-resolution” In Proc. of the IEEE International Conference on Computer Vision (ACCV), 2014, Darling Harbour, Sydney.

117. E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” In Proc. of the Computer Vision and Pattern Recognition (CVPR) Vol. 1, No. 2, p. 4, 2017, Hawai, USA.

118. M.S. Uddin, P.P. Das, and M.S.A. Roney, “Image-based approach for the detection of counterfeit banknote of Bangladesh,” In Proc. of the 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 1067-1072, 2016, Dhaka, Bangladesh.

119. B. Van Ginneken, C.M. Schaefer-Prokop, and M. Prokop, “Computer-aided diagnosis: how to move from the laboratory to the clinic,” In Radiology, Vol. 261, No. 3, pp. 719–732, 2011.

120. N.K. Verma, R. Abhishek, and S. Al, “An optimized fault diagnosis method for reciprocating air compressors based on SVM,” In Proc. of the IEEE International Conference on System Engineering and Technology (ICSET), pp 65-69, 2011, Shah Alam, Malaysia.

121. K. Verma, B. K. Singh, and A. Agarwal, “Indian currency recognition based on texture analysis,” In Proc. of the Nirma University International Conference in Engineering (NUiCONE), pp. 1-5, 2011, Gujarat, India.
122. R. Vidal, Y. Ma, and S.S, Sastry, “Robust Principal Component Analysis,” In Generalized Principal Component Analysis, Springer, pp. 63-122, 2016, New York, USA.
123. R. Vidhi, and S. Khushbu, “An integrated technique Dct & Sift for detection of image forgery,” In Int. J. Technol. Res. Eng., Vol. 1, No. 10, pp. 12-15, 2014.
124. Z.R. Wang, Y.L. Jia, H. Huang, and S.M. Tang, “Pedestrian detection using boosted HOG features,” In Proc. of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1155-1160, 2008, Beijing, China.
125. D. Wang, A. Khosla, R. Gargeya, H. Irshad and A.H. Beck, “Deep learning for identifying metastatic breast cancer,” Quantitative Methods, Cornel University, 2016, Available: arXiv preprint arXiv:1606.05718.
126. Z. Wang, D. Liu, J. Yang, W. Hannand T. Huang, “Deep networks for image super-resolution with sparse prior”, In Proc. of the IEEE International Conference on Computer Vision (ICCV), pp. 370–378, 2015, Las Condes, Chile.
127. T. Windeatt, R.S. Smith, and K. Dias, “Weighted Decoding ECOC for Facial Action Unit Classification,” In Proc. of the International Conference on Artificial Intelligence, pp. 26-30, 2008, Paris, France.
128. E. Xi, S. Bing, and Y. Jin, “Capsule network performance on complex data,” Machine Learning, Cornell University, 2017, arXiv preprint arXiv:1712.03480.
129. L. Xiaofeng, C. Tianhuang, and Q. Tingchao, “Making a virtual sand table based on unity 3D technique,” In Proc. of the 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), pp. 278-281, November 2014, Xian Ning, China.
130. Y. Yazici, K.H. Yap, and S. Winkler, “Autoregressive Generative Adversarial Networks” In Proc. of the Seventh International Conference on Learning Representations (ICLR), 2018, New Orleans, USA.
131. K. Yoshida, M. Kamruzzaman, F.A. Jewel, and R.F. Sajal, “Design and implementation of a machine vision based but low cost stand alone system for real time counterfeit

Bangladeshi banknote detection," In Proc. of the 10th international conference on Computer and information technology, pp. 1-5, 2007, Dhaka, Bangladesh.

132. X. You, R., Wang, and D. Tao, Diverse expected gradient active learning for relative attributes. In IEEE Transactions on Image Processing Vol. 23, No. 7, pp. 3203-3217, 2014.
133. P. P. Ypsilantis and G. Montana, "Learning what to look in chest x-rays with a recurrent visual attention model," Machine Learning, Cornell University, 2017, Available: arXiv preprint arXiv:1701.06452.
134. F. Zeggeye, and Y. Assabie, "Automatic Recognition and Counterfeit Detection of Ethiopian Paper Currency," In International Journal of Image, Graphics and Signal Processing, Vol. 8, No. 2, pp. 28-36, 2016.
135. K. Zhang, X. Gao, D. Tao and X. Li, "Multi-scale dictionary for single image super-resolution," In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1114-1121, 2012, Providence, Rhode Island.
136. R. Zhang, Y. Ming, and J. Sun, "Hand gesture recognition with SURF-BOF based on Gray threshold segmentation," In Proc. of the IEEE 13th International Conference on Signal Processing (ICSP), pp. 118-122, 2016, Chengdu, China.
137. Y. Zhang, H. Yan, X. Zou, F. Tao, and L. Zhang, "Image Threshold Processing Based on Simulated Annealing and Otsu Method" In Proc. of the 2015 Chinese Intelligent Systems Conference, Springer, pp. 223-231, 2016, Berlin Heidelberg, Germany.

Glossary

1. **Bidirectional Generative Adversarial Network**: Bidirectional Generative Adversarial Network learns the inverse mapping to project the data back to the latent space. It trains not only a generator but also an encoder that induces a distribution for mapping data point into the latent feature space of the generative model.
2. **Bag of Words**: Bag of Words is a simplifying representation used in natural language processing and information retrieval. A text or image is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.
3. **Convolutional Neural Network**: Convolutional Neural Network is a deep learning algorithm which can take in an input image, assign importance to various aspects/objects in the image and be able to differentiate one from the other.
4. **Computer-Aided Diagnosis**: Computer-Aided Diagnosis are systems that assist doctors in the interpretation of medical images.
5. **Deep Convolutional Generative Adversarial Network**: Deep Convolutional Generative Adversarial Network mainly composes of convolution layers without max pooling or fully connected layers. It uses convolutional stride and transposed convolution for the downsampling and the Upsampling.
6. **Generative Adversarial Network**: Generative Adversarial Network is a class of machine learning systems. Two neural networks contest with each other in a zero-sum game framework.
7. **Gradient Location and Orientation Histogram**: Gradient Location and Orientation Histogram is a robust image descriptor that can be used in computer vision tasks. It is a SIFT-like descriptor that considers more spatial regions for the histograms.
8. **Histogram of Gradient**: Histogram of Gradient is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.
9. **Inception Score**: Inception Score is a metric for automatically evaluating the quality of image generative models.

10. **Mean Square Error**: Mean Square Error of an estimator (of a procedure for estimating an unobserved quantity) measures the *average* of the squares of the errors; that is, the average squared difference between the estimated values and what is estimated.
11. **Multi-Layer Perceptron**: Multi-Layer Perceptron is a class of feedforward artificial neural network. It consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.
12. **One Versus All**: One Versus All involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. This strategy requires the base classifiers to produce a real-valued confidence score for its decision, rather than just a class label; discrete class labels alone can lead to ambiguities, where multiple classes are predicted for a single sample.
13. **One Versus One**: One Versus One train a binary classifier for a K-way multiclass problem; each receives the samples of a pair of classes from the original training set and must learn to distinguish these two classes. At prediction time, a voting scheme is applied: all classifiers are applied to an unseen sample and the class that got the highest number of predictions gets predicted by the combined classifier.
14. **Peak Signal to Noise Ratio**: Peak Signal to Noise Ratio is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.
15. **Principal Component Analysis**: Principal Component Analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables a set of values of linearly uncorrelated variables called principal components.
16. **Recurrent Neural Network**: Recurrent Neural Network is a class of artificial neural network where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior.
17. **Scale Invariant Feature Transform**: Scale Invariant Feature Transform is a feature detection algorithm in computer vision to detect and describe local features in images. SIFT keypoints of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature

from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of keypoints that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches.

18. **Speeded Up Robust Feature**: Speeded Up Robust Feature is a patented local feature detector and descriptor that uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image.
19. **Super Resolution Generative Adversarial Network**: Super Resolution Generative Adversarial Network is a deep learning approach for image super-resolution capable of inferring photo-realistic natural images for 4× upscaling factors.
20. **Super Resolution Convolutional Neural Network**: Super Resolution Convolutional Neural Network is a deep learning approach for single image super-resolution (SR) that learns an end-to-end mapping between low and high-resolution images, with little extra pre/post-processing beyond the optimization.
21. **Support Vector Machine**: Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used by building a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier for classification and regression analysis.
22. **Structural Similarity Index**: Structural Similarity Index is a method for predicting the perceived quality of digital images and videos.
23. **K-Nearest Neighbor**: K-Nearest Neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.
24. **Weighted Majority Average**: Weighted Majority Average is a heterogeneous ensemble learning algorithm employed in constructing a compound algorithm from a pool of prediction algorithms. WMA is a binary decision problem and builds a compound algorithm by assigning a positive weight to each base classifier.