PrFA: a framework for image-based counterfeit coin detection using Pruned Fuzzy Associative Classifier

Seyedeh Maryam Sharifi Rad

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 in Computer Science

Concordia University
Montreal, Quebec, Canada

October 2020

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Seyedeh Maryam Sharifi Rad

Entitled: PrFA: a framework for image-based counterfeit coin detection using pruned fuzzy associative classifier

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

Doctor Of Philosophy (Computer Science and Software Engineering)

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. Arash Mohammadi

______________________________ External Examiner
Dr. Benjamin Fung

______________________________ External to Program
Dr. Chunyan Wang

______________________________ Examiner
Dr. Adam Krzyzak

______________________________ Examiner
Dr. Juergen Rilling

______________________________ Thesis Co-Supervisor
Dr. Ching Y. Suen

Approved by ______________________________
Dr. Leila Kosseim, Graduate Program Director

October 15, 2020

____
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

PrFA: a framework for image-based counterfeit coin detection using Pruned Fuzzy Associative Classifier

Seyedeh Maryam Sharifi Rad, Ph. D.
Concordia University, 2020

The numismatic industry has been dealing with the problem of counterfeit coins for centuries. With the growth of technology in counterfeiting, untrained users cannot distinguish fake from genuine coins, especially for rare and precious coins. Using coin experts is also very expensive, and even with them, in some cases, detecting counterfeit coin is not guaranteed. Consequently, the demand for a computer-aided system that can detect counterfeit coins and be robust and reliable has increased. This thesis proposes a new fake coin detection method that focuses on image mining approaches instead of extracting only statistical features from the coin image, as suggested by other researchers.

In this research, a novel framework called *PrFA* is proposed for counterfeit coin detection that shows the effectiveness of image mining techniques. We develop a new image mining system on top of the fuzzy concept that helps us to discover the implicit information from the images in the way closer to the human's viewpoint. The advantage of the fuzzy set concept is that it can deal with uncertain objects, and we take this advantage for the decision-making problem by implementing an associative classifier model.

Our proposed framework is developed in two modules, and the principle of least privilege of it is a compressed system that can be considered as a knowledge attainment tool. In the first module, a method to detect the region of interests (ROIs) is applied that focuses on blob detection. In the second module, image mining is applied to find image patterns present in coin images using fuzzy association rules mining.

Image data are generally high dimensional due to a wide range of resolution levels. According to state of the art, the rule-based association methods demonstrated their efficiency by generating defensible solutions at an acceptable level of accuracy when dealing with small and medium-size samples. Regrettably, to cope with a large amount of data such as the image database, these methods were not robust enough. To tackle the above

challenge, we propose a new algorithm for feature selection to reduce the dimensions of features via analyzing the relationships among different features.

Image classification is imperative to search for more available and appropriate information. In recent years, various methods based on image mining approaches for classification tasks have been explored. Apart from their usefulness, the available classifiers are often vulnerable to low accuracy. Accordingly, we present a pruned based fuzzy associative classifier algorithm to create a robust counterfeit coin detector system. This classifier is a mixture of the association rules method and the fuzzy set concept.

In this research, we preserve the full power of fuzzy association rule mining to reduce the amount of redundant and insignificant rules by focusing on pruning methods. By comparing the achieved results with some other methods obtained from the same dataset, we demonstrate that our framework surpasses in terms of lower feature dimensions, and smoother boundaries while maintaining satisfactory accuracy. Besides that, we show that our proposed classifier is more accurate compared to other associative classifiers.

In this research, the problem with a general form will be described to provide a common framework for issues appearing in other domains.

Acknowledgments

I would like to express my gratitude to all those who gave me the possibility to complete this dissertation. First and foremost, thanks in large part to the kindness and considerable mentoring provided by Professor Ching Y. Suen, my dissertation advisor. His help, advice, and encouragement guided the way for this long journey. Prof. Suen was always available to openly discuss new research ideas and to take the time to review my research articles and thesis drafts. I want to offer my heartfelt thanks to him, who gave insightful comments and invaluable suggestions to improve this research. Throughout this research, I felt honoured to be working with a star professor in image processing and pattern recognition. Prof. C. Y. Suen is a philosopher and a guide with unconditional support, and he is the best supervisor that anyone would hope for.

I would also like to extend my gratitude to my examination committee for taking the time to read and evaluate this thesis. Their comments and feedbacks are valuable and highly appreciated.

I thank all my friends, colleagues, and staff at the Centre for Pattern Recognition and Machine Intelligence (CENPARMI) for their assistance and encouragement. Particular thanks go to Dr. Marleah Blom, CENPARMI's Executive Assistant, for her helpful administrative assistance. She is someone that I will always adore and admire. This work would not have been possible without the technical support of Mr. Nicola Nobile, CENPARMI's research manager. He deserves many thanks for the help and assistance during these past years.

Nobody has been more important to me in the pursuit of this project than my beloved husband, Saeed Khazaee. I would like to express my utmost gratitude to him, whose love and guidance are with me in whatever I pursue.

Last but not least, I want to express my heartfelt gratitude and appreciation to my parents, who believed in me by all means throughout my journey. Their unlimited supports made everything possible to help me make it this far.

Declaration

The work described in this thesis has not been previously submitted for a degree in this or any other university, and unless otherwise referenced, it is the author's work. No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Maryam Sharifi Rad

Concordia University

Oct. 2020

Statement of Copyright

Publications

Publications produced towards Ph.D. candidature:

a) Publications generated directly from doctoral research:

1. S. Khazaee**, M. Sharifi Rad**, C. Y. Suen, “Detecting of Counterfeit Coins Based on Modeling and Restoration of 3D Images”, *In Barneva R., Brimkov V., Tavares J. (eds) Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications. Lecture Notes in Computer Science, Springer,* Volume 10149, 2017, pp. 178-193.
2. **Sharifi M**., Khazaee, S**.,** Suen, C., “Counterfeit Coin Detection Based on Image Content by Fuzzy Association Rules Mining”, *In proceeding of ICPRAI 2018, Montreal, Canada, Center for Pattern Recognition and Machine Intelligence*, 2018, p. 285-289.
3. **M. Sharifi Rad**, S. Khazaee, L. Liu, and C.Y. Suen, “A Blob Detector Images-Based Method for Counterfeit Coin Detection by Fuzzy Association Rules Mining”, *International Conference on Pattern Recognition and Artificial Intelligence, Springer, Cham, 2020*, pp. 669-684.
4. **M. Sharifi Rad**, S. Khazaee, L. Liu, and C.Y. Suen, “A framework for image-based counterfeit coin detection using pruned fuzzy associative classifier” *International Journal on Wavelets, Multiresolution, and Information Processing (IJWMIP), In press*.
5. **M. Sharifi Rad**, S. Khazaee, L. Liu, and C.Y. Suen, “A Pruned Fuzzy Associative Classifier to Detect Image-Based Counterfeit Coins”, Pattern Analysis and Applications, (Under Minor revision)

b) Publications generated from research collaborations in related topics:

6. Khazaee, S., **Sharifi M.**, Suen, C., “Restoring heightmap images of shiny coins using spline approximation to detect counterfeit coins”, In proceeding of ICPRAI 2018, Montreal, Canada, Center for Pattern Recognition and Machine Intelligence, 2018, pp. 383-387.
7. Saeed Khazaee, **Maryam Sharifi Rad**, Ching. Y. Suen, "Detection of counterfeit coins based on 3D Height-Map Image Analysis ", *Expert Systems with Applications*, (*Under Minor revision*)
8. Saeed Khazaee, **Maryam Sharifi Rad**, Ching. Y. Suen, “Decomposing relief maps to detect counterfeit coins using a hybrid deep learning approach”, *Journal of Visual Communication and Image Representation*, (*Under review*)

To my love …

Saeed Khazaee;

Who made it all possible, who taught me to smile in front of hardness, to value myself when lightning strikes me, to fight for my rights, to chase my dreams, and to appreciate blessings of life.

To my love when he always tells me
the one who walks on the right track would
ultimately arrive…

Table of Contents

List of Figures

List of Tables

ROI	**R**egion **o**f **I**nterest
DM	**D**ata **M**ining
KDD	**K**nowledge **D**iscovery in **D**ata
AR	**A**ssociation **R**ule
ARM	**A**ssociation **R**ule **M**ining
CI	**C**omputational **I**ntelligence
AI	**A**rtificial **I**ntelligence
SC	**S**oft **C**omputing
ANN	**A**rtificial **N**eural **N**etworks
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
SURF	**S**peeded **U**p **R**obust **F**eatures
EA	**E**volutionary **A**lgorithm
GA	**G**enetic **A**lgorithm
PSO	**P**article **S**warm **O**ptimization
FGBRMA	**F**uzzy **G**rid-**B**ased **R**ule **M**ining **A**lgorithm
KNN	**K**-**N**earest **N**eighbor
CNN	**C**onvolutional **N**eural **N**etwork

Chapter 1

Introduction

This document contains my thesis research on the problem of detecting counterfeit coins. We tackle this problem in the context of developing a general framework that shows the effectiveness of image mining techniques. Counterfeit coin detection is a classic research topic, and it has attracted much interest in other fields such as numismatics and forensic investigation. This chapter briefly describes the background to the problems investigated in this thesis, the motivations of the work, major contributions and the structure of the thesis. In this chapter, the problem statement and related research issues will be discussed, and throughout the document, a general framework to detect fake coins will be proposed.

1.1 *Problem Statement*

With the fast evolution of multimedia technologies and intelligent systems, image processing and pattern recognition have become a hot topic among studies. On the other hand, with the growth of data and a huge number of digital photographs, data mining, and knowledge discovery have become a priority. Nowadays, image mining, which can automatically review meaningful information from image datasets, has established its position as a prominent and essential research area. For the past few years, a huge number of digital images have been generated on multimedia applications. As a result, building a computer-aided system to mine valuable information from the extensive digital image datasets is becoming a priority for many researchers. With the consideration of this trend and the vast amount of data involved in all kinds of digital images, image mining approaches seem to be appropriate for this purpose. This trend has affected our daily life in every way, and advances in image acquisition have led to tremendous growth in significantly large image datasets. Image mining is an interdisciplinary endeavour that draws upon techniques developed in computer vision, image processing, image retrieval, machine learning, and artificial intelligence, etc.[2].

In recent years, a lot of illegal counterfeiting rings manufacture and sell fake coins, which have caused great loss and damage to society [3]. On the other hand, collecting ancient and precious coins has become fascinating for many people around the world. In many cases, ancient and rare coins have also been traded pricing for more than one million dollars. Because it evokes prompt and relative connectivity to the ancient past, one that we can literally hold in our hand, the collecting of ancient coins has become an almost addictive pastime for many people today. By massive progress in technology and science and interdisciplinary approaches in all academic majors, numismatics is essential for our understanding of the ancient thoughts and lifestyles [4]. In the past few years, after two major workshop competitions in 2006 and 2007 by MUSCLE CIS-Benchmark, counterfeit coin detection has received enormous attention. The competition aimed to classify European coins belonging to 12 countries before the introduction of the euro coins. Counterfeit coin detection is a complicated process because of widely varying input patterns, cluttered images, and various rotations, which are enormous challenges. As an essential topic of security, counterfeit coin detection has become the focus of research in the field of numismatics. During the past fifteen years, the digital investigation has earned much more consideration despite being a comparatively nascent scientific field [5]. In recent years, various methods have been proposed based on the physical properties of coins such as thickness, size, weight, diameter, conductivity, metallic properties, properties of the electromagnetic field. Although forensic experts can be employed to detect the suspected coins, having a human label every possible object in a vast collection of coin images is a daunting task. Accordingly, an automatic counterfeit coin detection system can converge toward the desired point dramatically. Besides, computer-aided systems are much more efficient, require no human expertise, and most are portable. There is a growing interest in counterfeit coin detection community toward the application of image mining techniques in this field. Digital image forensics is used to mine the critical information related to digital images, and mining of this important information facilitates realizing the legitimacy and truthfulness of digital images [6]. Although several studies have been developed based on different approaches for mining of images to extract strong association rules, yet it is a challenging task. Most of the existing associative classification algorithms relied only on *Support* and *Confidence* parameters, failing to equip some statistical metrics for rule evaluation, and often suffer from low accurate or complicatedly interpretable results to decision-makers [2], [7]–[11]. Furthermore, in a counterfeit coin detection application, it is unrealistic to present a huge number of rules to forensic experts for further studies and manual tuning.

Existing studies have demonstrated that most of the extracted rules are redundant or irrelevant. Despite the efficiency of association rules mining, this strength comes with a major drawback. What makes it difficult is to analyze the huge number of rules to identify those interesting/useful ones [12]. The question is: "Can we preserve the full power of fuzzy association rule mining (i.e., its completeness) for counterfeit coin detection without overwhelming the user?"

Based on the type of analysis carried out [13]–[18], the potential of a fuzzy associative classifier has not been applied to coin image datasets for counterfeit coin detection and is still an open field of research. It is worthy to note that fuzzy logic can be used to identify complex patterns or structural variations in image datasets, and the combination of association rules mining and fuzzy sets theory presents a new approach for better counterfeit coin detection with flexible prediction power.

In this research, a novel framework called *PrFA* based on fuzzy concepts to rely on semi-automatic image mining for counterfeit coin detection will be proposed. We design a pruned fuzzy associative classifier which has an optimization procedure of search for image features and strong fuzzy rules that better satisfy the desired measures for counterfeit coin detector system. To obtain an accurate fuzzy associative classifier, we apply two post-processed pruning techniques based on redundancy restriction and feature dimensionality reduction. To do this, a new engine based on a fuzzy association feature selection method will be proposed, which is incorporated into the proposed framework to control the size of the feature subset of images. As a result of this, the feasibility of applying fuzzy association rules mining to detect counterfeit coins will be demonstrated. Even though some methods have been introduced in the literature for counterfeit coin detection, we believe that this is the first work that addresses these problems in the context of fuzzy association rules mining based on the image content.

The Danish law enforcement department initiated this research started the investigations in collaboration with forensic technology companies in Montreal. The Law Enforcement Office provided the coins containing both genuine and counterfeit samples used in this research. It is important to note that having access to fake coins is an arduous undertaking due to the legal issue, and the access to more fake coins is usually restricted. The dataset applied in this research consists of real data, and we scanned the coins by a powerful 3-D scanner that could produce both 3-D and 2-D images. The machine that was used for capturing the coins was a powerful scanning device named IBIS TRAX patented by Ultra Electronics Forensic Technology Ltd. Co in Montreal, Canada [19]. IBIS TRAX-3D is a scanning device that can digitize a coin in 2D (with fast capturing) and 3D (with slow capturing). This scanning device enlarges the original image and presents delicate details into high-resolution views.

1.2 *The Aim and Objectives*

The image mining concept deals with implicit knowledge extraction, image data relationship and other patterns that are not clearly stored in the images. Most studies concerning image datasets focus basically on low-level features and neglect the conceptual association among the objects in the images. Based on the type of analysis carried out, several studies have proposed the methods for mining association rules in relational datasets. In these cases, the data is explicit, and there is a specific item for each transaction. It is noted that these methods are not suitable to extract rules in the image datasets because there are essential differences between relational datasets and image datasets. Here we summarize the two main differences between relational and image datasets [2]:

1. Absolute versus relative values. In relational datasets, the data values are semantically meaningful. However, in image datasets, the data values themselves may not be significant unless the domain expert supports them.
2. Spatial information (Independent versus dependent positions). The implicit spatial information is critical for the interpretation of image contents, but there is no such requirement in relational datasets.

As can be seen, there is a major deviation from the typical data mining approach. In most real-world applications, concepts often relate to each other instead of appearing in isolation.

Although several methods have been applied to extract association rules, many of these methods have failed to manage a large number of rules. A large amount of data is often extracted from a single image without further details about the rules, which makes it difficult for knowledge interpreters to evaluate the rules. To deal with this domain, some researchers [2], [7]–[11], [20]–[26] have developed mechanisms that relied only on Support and Confidence parameters, while others preferred to apply a method to analyze the huge number of rules [27]–[30].

This thesis introduces a solution to the problem where the major question is how to create an accurate counterfeit coin detection system without the complexity. In contrast, existing studies have demonstrated that most of the extracted association rules are redundant or irrelevant. The dissertation seeks an alternative method to rely on semi-automatic mining on the image datasets. From the data mining techniques usually used in similar domains, association rules mining is chosen because it has been shown to work remarkably well in discovering the implicit information from the images.

On the other hand, there is growing evidence that integrating association rules mining and classification develops more efficient and accurate classifiers than traditional methods. This research is broadly oriented towards developing a framework based on fuzzy concepts to rely on semi-automatic image mining for counterfeit coin detection that achieves promising classification performance.

It is a first attempt to combine fuzzy association rules and the coin image dataset, although there has been remarkable research in the image domain. In this research, a method to discover the frequent objects and to extract the interesting fuzzy association rules for counterfeit coin detection will be proposed. The fundamental component in image mining is to identify similar objects in different images. We transform the images with a set of transactions, each transaction representing one image with the features extracted as well as other given characteristics along with the class label. The result of this phase is a transactional dataset to be mined in the next phase of our proposed framework. In this way, the unclassified images can be automatically classified by generated fuzzy association rules.

From another point of view, the most delicate part of the classification with fuzzy association rules mining is the construction of the classifier itself. Thus, the objective of this task is to develop a technique that performs well in real-world domains while being computationally efficient. The dissertation will describe a framework addressing all the above issues efficiently and effectively.

This thesis aims to provide a common framework in scenarios that require image classification. It tries to solve the problem of image-based counterfeit coin detection by fuzzy association rules mining. Discovering useful knowledge is a challenging endeavour. Many of the issues for classification we face are in other areas as well. Accordingly, this study tries to separate the problems that we have the most interest in the database community in the field of image mining research. Our proposed framework will allow us to discover interesting rules among the massive amount of trivial and diffused associations and will demonstrate the feasibility of future work in this research line. In particular, it is focused on extracting individually significant rules and avoiding the exponential curse [31] that would make the subsequent expert examination unappealing.

Background preliminary concepts about these problems are introduced in the coming chapters.

1.3 *Challenges*

There are many interesting challenges to be explored in the domain. The main obstacles of this research can be summarized below:

- Although coins are made of alloys and are more durable than banknotes, yet they are sensitive to corrosion and rust. Because of their low-value, new coins by long intervals replace the old ones. This challenge is more prominent when we want to process the image of an ancient coin since they can be changed in some edges around the shapes or letters. As we do not concentrate on the edge-based features in this research, the problem will have a lesser impact on the method, however, damaged coins still remain a challenge for this research.
- A comprehensive method should be able to recognize any type of coins for recognition. Since coins differ not only in size, characters, and shapes but also in the language of the country of origin, this would be very challenging to face these variations.
- Illumination variation of shiny coins adversely affects the process of image-based counterfeit coin detection that must be considered during the scanning process.
- Imbalanced classes are a common challenge in any classification problem where there is an inappropriate ratio of observations in each class. Accessing counterfeit coins is highly challenging in counterfeit coin detection research. Law enforcement departments do not release fake coins or their images for security reasons. However, to compensate for the lack of enough fake samples for this research, we take advantage of Deep Learning to produce fake samples. We use the Generative Adversarial Neural Network to generate counterfeit coin images and augment the fake class data.
- The creation of a suitable format of the fuzzy image dataset is a more prominent challenge when it needs to be mined for feature extraction. In this research, we model the images with a set of transactions, such that each transaction represents one image with the visual blobs extracted along with the class label.

- The association rules mining always results in a large number of rules with the growing size of datasets, which can cause computational complexity and increase the risk of overfitting. This is one of the main challenges for the proposed framework. To address this issue, a pruning method will be presented to reduce the number of rules and keep the most important rules.

1.4 *Major Contributions*

The main contributions of this research are:

- Creation of datasets. We create six coin image datasets containing genuine and fake coin data, which can be used for training and evaluation of fake coin detection approaches.
- Introduction of the *PrFA* framework. We develop an image mining system on top of the concept of fuzzy association rules mining using a blob detection method and relationship predicate that helps us to discover the implicit information from the images in a way closer to the human's viewpoint.
- Proposing a new feature selection method. We propose a new engine based on a fuzzy association feature selection method, which incorporated in our framework to reduce the size of the feature subset of images and improve the computational performance of our framework.
- Designing a fuzzy associative classifier. We design a pruned fuzzy associative classifier that has an optimization procedure of search for image features and strong fuzzy rules that better satisfy the desired measures for the counterfeit coin detector system.
- Performed an experimental evaluation and comparison of our *PrFA* framework with other image-based coin detection approaches.

1.5 *Thesis Organization*

The remainder of this thesis is organized as follows:

In chapter 2, a literature review of related research in the field of coin authentication and the existing research studies on counterfeit coin detection are discussed. Chapter 3 provides the state-of-the-art about the fundamental and essential concepts of association rule mining and its existing algorithms, and fuzzy logic theory, as well. In chapter 4, the proposed framework for counterfeit coin detection is demonstrated. It introduces evolutionary post-processing of pruning techniques to extract more useful and significant rules for classification. In Chapter 5, the experimental results are given to show the effectiveness of the proposed method. The experimental results show the superiority of this method against other approaches existing in the literature. Finally, the study is concluded with a summary of the primary contributions of this work and an outline of future work in chapter 6.

Chapter 2

Literature Review

In this chapter, we will thoroughly investigate the fundamental concepts and the overall objective of coin authentication. The existing research studies on counterfeit coin detection are discussed in this chapter.

2.1 *Introduction*

Coins are an integral part of our day to day life. They are used almost everywhere, such as supermarkets, grocery stores, banks, and other domains. Coin hoarding often involves thousands of coins, so coin classification is a time-consuming and cumbersome task, which up until now has to be carried out manually. Discovery in this field of study is attractive and meaningful because it supports further crime reduction. Coin forging is a prompt self-evolving industry. Due to the much-improved features of spurious coins made nowadays, a more detailed investigation is needed to identify counterfeit coins.

Although forensic experts can be employed to detect the suspected coins, it is evident that having a human label every possible object in a vast collection of coin images is a daunting task. That is where numismatists can make a profit from computer-aided methods to identify spurious coins. So, there is substantial demand for highly accurate and efficient automatic counterfeit coin detection systems [32]. Despite daily uses, fake coin detection systems can also be applied for research purposes by the institutes or organizations that deal with the ancient coins. In addition to highly precious collectable and antique coins, there are large numbers of counterfeit coins in circulation around the world, and that number is increasing per annum.

Recently, Canadian researcher, Mike Marshall, who is a well-known counterfeit coin educator and began his campaign against counterfeit coins in 2007, notified Canadian authorities about fake collectable Canadian coins available for sale on eBay's website [33]. According to his statements, online markets like eBay, Alibaba, DHgate, and all prominent e-commerce companies which each with numerous vendors offering counterfeit coins, have been inundated with fake coins. Mike Marshall believes that coin counterfeiting has been a constant problem since 2007 until now, and all kinds of counterfeit coins are still being spread every day in bourse and online markets. Fourteen million counterfeit Pound coins have been reported in the UK circulated between 2003 and 2004, while reports show more than 47 million fake £1 coins in 2014 as the Royal Mint estimation [32]. Fake coins have always been a major concern to law enforcement authorities throughout the years. Accordingly, governments are searching for reliable solutions to overcome the counterfeiting problem.

As mentioned earlier, the Danish law enforcement department initiated this research started the investigations in collaboration with the forensic technology companies in Montreal, Canada. In this research, the proposed framework is evaluated on a real-life dataset of Danish coins as part of collaborative research with Danish authorities.

2.2 *Related Works*

Traditionally, fake coins have been discovered by human experts, and identification was based on their expertise. The traditional manner of authenticating the coins includes manual methods, periodical investigation on coin dealers and sales catalogues, followed by a human investigation. Due to improvements in technologies, these methods have shown a lack of precision.

The most common method to forge a coin is to strike a coin using a fake die that is moulded from the original coin stamp. This method results in a small variation in salient width which is reflected as the edges in a coin image.

Several scientific methods may give experts a clue if a coin is counterfeit or not. The first is to have access to detailed specifications of a genuine coin. These should include size, diameter, thickness, metal composition, weight, and specific gravity. Areas they usually study include the shape of the letters, the position of numbers, details on portraits, and the overall look and feel of a genuine coin. Apart from the weight and the material of the coins, experts normally concentrate on the edges and local features of the coin with a magnifier. Most of them compare the coins with the original characteristics that are provided in a book or a magazine.

The experts usually know the weight of an original contemporary coin from a coin catalogue. Therefore, a counterfeit coin can be detected by a regular weight scale. Even if the scale is not very precise, it will show a big difference to the catalogue weight of the coin. Furthermore, the edge of a counterfeit coin is usually plain because the reeding or edge inscriptions are hard to copy. The experts usually know how the edge should look like and since most counterfeiters do not pay much attention to the details of a coin, the experts use a magnifying glass and can easily sort out most of the common fakes.

Some high-quality counterfeit coins have even fooled expert numismatists. It is important to understand the minting process for the individual coin that they are inspecting.

Based on experience, experts can identify some coins as counterfeit by touching and feeling the surface. In this process, they usually look for the primary features such as the depth of lettering, the sharpness of design, vagueness (fineness) of milled edge, etc. Unlike these features, relative position is an inter-lettering feature, which records the distance between every pair of adjacent letters and digits. Physically, these distances are millimeters long, and the difference over distance is even smaller to examine.

Apart from the coins' varying qualities, maturing counterfeiting technologies are narrowing the precision gap between fake and real coins, raising the average quality level very high. In response, however, associated knowledge did not evolve much. This unbalanced development results in more hassles for the general public and raises the need for more recognition research.

This is also an important concern in the field of numismatics. Forensic experts may be employed to examine the suspected coins, yet it is unrealistic considering the large quantities of coins that have to be examined. Therefore, an automatic fake coin detection system is highly desired.

In general, coin recognition systems are proposed for two types of coins: modern coins and ancient coins. Figure 2-1 demonstrates examples of ancient and modern coins. There are scant differences between the two types where the ancient coins have some shape abnormality due to the manual minting process. Few authors suggest that recognizing ancient coins are somewhat more challenging than modern coins due to the minting process, where coins of the same class can vary based on the minting master [34]. Nevertheless, research articles apply the same feature extraction and classification methods for both modern and ancient coins.

(a) (b)

Figure 2-1 Examples of ancient and modern coin images: (a) ancient Roman Republican coin, and (b) Canadian two-dollar coin.

There are different types of counterfeit coin detection systems, including Mechanical, Electromagnetic, and Image processing based systems available in the market [35]. The mechanical systems utilize parameters like radius, diameter, weight, thickness, shape and so on to distinguish the coins. Some studies have utilized sound and light to authenticate coins such that they applied frequencies captured to recognize spurious from genuine coins. Nevertheless, these frequencies depend mostly on the nature of metal. Such that, if the identical metal is exploited for the counterfeit coin, it will pass the test. Although these measurements may be used to distinguish between genuine and fake coins as presented in the patents, they are somewhat limited in the discriminative power. The counterfeit coins manufactured nowadays are of high quality, so they often comprise primary likeness to their genuine counterparts, which causes the detection of fake coins extremely challenging. However, these parameters cannot be applied to identify the differences between the various materials of the coins. Alternatively, when the physical properties of the coins are very similar, image-based counterfeit coin detection systems can be used.

In electromagnetic systems, X-ray fluorescence is used to authenticate coins. X-ray is naturally detrimental, and the utilization of fluorescence X-ray is an infeasible and expensive process. In these systems, the coins are moved through an oscillating coil at a specific frequency. For example, in the patent [36], counterfeit coins are detected by an electromagnetic method-based system. In this method, an oscillation coil excited by an exciting signal containing a plurality of harmonic components is passed on one side of a coin. For the other side of the coin, it is connected with a secondary oscillation coil that is coupled with the primary oscillation coil. When a coin passes through, the voltage signal coming from the secondary oscillation coil can be applied to specify whether the coin is original or fake. Also, an electromagnetic method-based system has for fake coin detection using the magnetic properties of the coin has been proposed in the patent [37]. Furthermore, authors in [38], [39] utilize X-ray fluorescence (XRF) for quantitative analysis by a metallic microscope for observation of coin microstructures to detect fake coins. These systems can provide the distinction among various materials and the accuracy of detection, but still, they can be fooled by some game coins.

Therefore, the application of image processing for counterfeit coin detection is more accurate and promising. In recent years, coins have been obtaining more attention from intelligence researchers, and image-based counterfeit coin detection systems have also come into the picture. Exploiting images to detect counterfeit coins usually benefits from low cost and high facility. With the growing popularity of smartphones, it is a desirable idea to determine the accuracy of a coin by taking a picture of it. In Image-based counterfeit coin detection systems, first, the images of the coins (obverse and back sides) are taken either by camera or by the scanning device and stored in a database along with textual descriptions. Then these images are processed by using different techniques of image processing like FFT [13], [14], Gabor Wavelets [15], image subtraction [16], edge detection, segmentation, and decision trees [17]. Accordingly, various features are extracted from the images and then, image processing techniques and machine learning are applied to classify the set of features and recognize coins. In particular, depending on the image taken of a coin, some pattern recognition techniques can be used to authenticate whether it is real or fake automatically.

Although a few algorithms have been conducted in the literature on coin authentication with promising results, it is still immature, and the demands of a reliable and efficient method are increasing. In recent years, most of the computer-aided methods are designed for coin recognition [40]–[44], while there are a few attempts in the literature that exploit images for counterfeit coin detection. Coin recognition systems usually utilize different image processing techniques for feature extraction and decision making to decide where each coin belongs. Machine learning techniques are usually applied to classify the set of features and recognize coins. Feature extraction is an essential step in coin detection systems, and the features are extracted based on various techniques like local image features [45], [46], texture features [47], [48], edge-based statistical features [49], SIFT [45], [46], and SURF [46], [50]. For example, [24] were among the first to present an end-to-end recognition workflow for ancient coins. They experimented with different interest point detectors and local image descriptors in order to specify the best combination. Recognition is accomplished by discovering the nearest neighbour using the Euclidean distance. The authors in [51] extended the proposed method in [24] by introducing an additional step which exploits the coin's contour as a characteristic feature for single coin samples. The contour is analyzed by intersecting rays cast from the coin's center of gravity with the coin border. The distances between the intersection points and a possible perfect circle fit to the coin contour are measured and used to build a descriptor, which can be computed quickly and allows for an efficient preselection. In [52], an algorithm for coin recognition using

Circular Hough Transform (CHT) has been proposed. The proposed system first uses canny edge detection to generate an edge map, then uses CHT to recognize the coins and further find their radius. The basic limitation arises when the image is captured from a distance. In [53], a system for Indian coin recognition by Heuristic approach and Hough Transform (HT) has been proposed. The proposed method has been limited to recognizing only the Indian Coins. Furthermore, there are some methods for coin classification in which edge information has been used as features of images with Fourier transform [54], [55]. Nevertheless, edge features are not strong enough because they are undeniably distorted by noises such as rust, dust, and abrasion. In addition to the above, there are several rotation-invariant local binary pattern approaches, which divide a coin image into several rings and then compute the histogram for each ring [56], [57]. These approaches, however, come with a major drawback. They generate a high dimensional feature and lead to a longer running time along with poor performance.

In [58], the authors have presented a method for the coin classification. Their system can distinguish 500 yen from 500 won coins. After separating the coin from the background, it is divided into several loops by a polar log network. This separation ensures the rotation is guaranteed. For each section, the number of gray surfaces in a multi-layered neural network is calculated and fed. The authors in [59] proposed a method for the recognition of 1-, 2- and 5-rupee coins. After successfully identifying the edge, the location of the numbers depicted on the coin is found by matching the template. In the next step, Gabor features are computed for the sub-image containing the numeral, which is then classified by a backpropagation network. In [17], a classification system for coin recognition has been introduced. This approach first segments the coin from the background using a generalized Hough transform. Next, the segmented image is converted to polar coordinates and gradients are computed. The gradients are quantized into different orientation bins and for each of which a binary image is constructed. The binary images are applied as feature vectors and are classified using the nearest neighbour classification.

Most of the methods outlined above were presented to identify and classify coins into political history, religion, and country minting. However, coin recognition remarkably differs from counterfeit coin detection investigated in this research. One of the main concerns in coin recognition is to decrease sensitivity to changes in coins from the same class. Therefore, it is very likely that counterfeit coins, especially those of high quality, will be classified in a category belonging to their original counterparts in the context of the coin recognition system. However, these specific changes may be a very beneficial sign that demonstrates the validity of the coin, so it can be underlined to identify counterfeit coins. Besides, it is more difficult to detect counterfeit coins because of the possibility that the above changes may be due to coin wear or contamination from daily use [3]. Although coin recognition is one of the areas of research that has attracted high interest and effort, not much research has been conducted on counterfeit coin detection as a close topic. Therefore, in this research, we focus on counterfeit coin detection to fill in the research gap.

Generally, a counterfeit coin is a simulation of a genuine coin manufactured with the purpose of cheating. Over the past years, a few studies have just been conducted to identify fake coins through different denominations such as Artificial Neural Networks, Circular Hough Transform, and Heuristics. Up to now, a few studies based on pattern recognition techniques and classification algorithms have been proposed that exploit images for counterfeit coin detection. In order to take advantage of images to detect counterfeit coins, the first step is image representation. In comparison with utilizing raw pixels for image representation, the feature extraction method that characterizes several different aspects of the image, such as its texture or shape, is a preferred way. Generally, a popular way is to put these features into a vector for image representation. Image vectorial representation facilitates access to machine learning tools such as artificial neural networks [60], [61] and SVM [62], [63] that work in the vector space. Although vectorial representation offers compactness, it lacks adequate descriptive power.

In addition to the vectors, some of the studies [64]–[67] presented various ways of image representation like trees and graphs. Furthermore, in [68]–[71], the local keypoint detectors and descriptors have been utilized for image representation. Although a set of keypoints can describe an image, they cannot assist the access to the vector-based machine learning tools.

Some studies have used images for counterfeit coin detection stand-alone or in conjunction with the methods mentioned above. Related work to detect fake coins, we can mention the following.

In [3], an image-based approach to detect the fake coins based on the characteristics of coin images has been proposed. The number of prototypes determines the dimension of coin images. They have computed the dissimilarity between the coin images by the local key points on each image using the DOG detector and the SIFT descriptor. The DOG detector is applied to find the key points which are defined using the SIFT descriptor. Each comparison between the test image and the predefined image is stored as a vector in dissimilarity space. Finally, the SVM is utilized to classify the coins into a genuine or fake class. In [14], a method to detect two-euro fake coins based on the coin images digitized by an optical mouse sensor has been developed. In this research, the basic limitation arises during the coin rotation, and it is vulnerable to distortions. In [15], a counterfeit coin detection method to detect fake Danish coins based on their image characteristics has been proposed. Despite the promising results achieved, the dataset used was extremely small, which consisted of only 16 coins. Due to the minimal experimental dataset, this method claims no guarantee to fit other coins. In [16], a counterfeit coin detection method based on 3D images was proposed. The authors suggested a straightening algorithm to convert each circular coin image to a linear rectangular image. Such that, the outer circle of the coin where all characters and numbers occur using the height and depth information obtained by the 3D scanner to identify genuine coins. The method has demonstrated promising results when images from specialized 3D scanner are digitized. Access to such a scanner is not conceivable in daily life, and it needs the expertise to digitize 3D images of coins. In papers [17], [18], counterfeit banknote detection systems based on image content have been proposed.

Several image-based coin classification systems have been developed in the past. However, the development of an image-based counterfeit coin detection system based on image mining techniques is still subject to ongoing research.

2.3 *Image Processing*

Today, image processing is one of the fastest-growing technologies. It is also a significant area of research in the fields of engineering and computer science disciplines too. Image processing is a method to execute some operations on an image in order to achieve an enhanced image or to extract some sufficient information from it. Image processing includes the following three steps:

- Image import through image access tools
- Image analysis and manipulation
- The output in which the result can be changed is the image or report, which is based on image analysis.

Image segmentation is one of the image processing methods which partitions the image into regions that correspond to structural units in the scene or specifying objects of interest. It deals with defining the contour of the image and the region belonging to the background. It divides the image into two areas: the region depicting the image and the region belonging to the background. The first step of the classification process requires the segmentation of the image from the background of an image. In this application, the background is the dark zone outside the rim of the coin. Figure 2-2 demonstrates a typical coin image sampled from IBIS TRAX-3D. Figure 2-2 (a) includes a 2D grayscale image that possesses texture information; Figure 2-2 (b) is the corresponding depth image, meaning coin depth information is coded and illustrated in intensity value. The grayscale and depth image correspond to each other, for any pair of coin images. Obviously, manipulations on one image are transferable to the other. The background does not supply beneficial information, yet it wastes RAM and pulls down calculation speed. Due to the complexity of the background and the colour variance between the coin and the background, an ordinary threshold operation cannot be appropriate for coin segmentation. Recent research in coin classification suggests two different segmentation approaches:

- Edge-based segmentation: This category of segmentation methods partitions an image based on abrupt changes in the intensity, i.e. edges found in an image by edge detectors.
- Hough transformation: This category of segmentation methods is used to isolate features of a particular shape within an image.

In [52], the authors have utilized edge-based segmentation and Hough transformation methods for ancient coin recognition. They have argued the capability of the edge-based method to segment ancient coins but not the Hough transformation. The authors reported that it is due to the unusual shapes of ancient coins, which tend to be, but are not entirely circular. However, for modern coins, the Hough transform method performed better than the edge-based method. Besides, a few other segmentation methods are based on an active model. Nevertheless, these methods are not as accurate as other coin segmentation methods in the literature.

(a) (b)

Figure 2-2 Image example from IBIS-TRAX-3D: (a) 2D gray-scale image, and (b) depth image.

2.4 *Segmentation of Digital Images*

The segmentation of images to find a region of interest (ROIs) is one of the most critical processing steps in image processing. By extracting ROIs that roughly correspond to the objects, we access the images at the level of objects rather than global image properties. BLOB stands for Binary Large Objects. Blob detection [72] is the focus of the first module of our research. It is applied to represent a group of pixels having similar values for intensity but different from the ones surrounding it. Blobs are usually characterized by a quite homogeneous interior and are surrounded by a boundary edge. In image processing, blobs (particles or dots) are considered as small structures whose visual properties, e.g. brightness or color, are different from those in their surrounding region. Blob detection is a transformation method from the raw pixel data to a small set of localized coherent regions in color and textual space. In this method, the pixels will be grouped into regions by modelling the joint distribution of color, texture, and position features. As the input is a binary image with the focus of interest valued one and background zero, the idea is to traverse all pixels by moving row by row. An image is made up of a collection of blobs, and each blob represents a region of the image which is relatively homogeneous in colour and texture. Each "blob" is considered as a 2-D ellipse which possesses several attributes. The pixels will be grouped into regions by modelling the joint distribution of colour, texture, and position features. Creating the blob representation of an image involves three steps:

1. Select an appropriate scale for each pixel and extract colour, texture, and position features for that pixel at the selected scale.
2. Group pixels into regions by modelling the distribution of pixel features with a mixture of Gaussians using the Expectation-Maximization algorithm.
3. Describe the colour distribution and texture of each region for use in a query.

Figure 2-3 [72] illustrates these steps for a sample image from pixels to region descriptions.

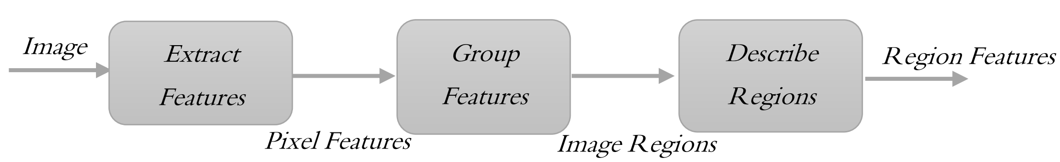


Figure 2-3 Stage of blob extraction.

After the above process, all objects that are detected can be called blobs. An image is made up of a collection of blobs, and each blob represents a region of the image, which is relatively homogeneous based on colour, texture, and spatial descriptors.

2.5 *Summary*

In this chapter, the preliminary concepts and the overall objective of coin authentication have been thoroughly introduced. Besides, the existing research studies on counterfeit coin detection were discussed. In this chapter, we tried to review image processing and image segmentation.

Chapter 3

Association Rules Mining

In this chapter, the overall objective of image mining, fuzzy logic theory, related research on associative, propositional, and structural approaches to classification will be introduced.

3.1 *Introduction*

Data mining is an effective technique for extracting useful knowledge from data sources. Association Rule Mining (ARM) is one of the significant tasks of data mining. ARM techniques have been successfully exploited in diverse fields such as market analysis, healthcare, industry, and recommendation systems [73].

ARM aims to discover close relationships among items in large datasets, which was popularized in particular due to the research by Agrawal et al. [74]. The research interest in association rule mining took off in 1993 with the presentation of Apriori, the most popular algorithm to extract association rules from a database. Since the presentation of the association rules mining technique, this area remained one of the current hot research topics in machine learning and knowledge discovery. ARM is a vital tool that has been applied in different industries like market basket analysis, fraud detection, forensic investigation, and several other applications where searching regularities are targeted. It determines interesting relationships among items in datasets. It can also be mentioned that the association rules mining technique is an appropriate tool for pattern detection. Nowadays, there is a growing interest in the application of image mining techniques, and in the last two decades, many researchers have presented the application of the data mining algorithms in diverse domains [1]. Image mining is an interdisciplinary endeavour that draws upon expertise in data mining, database, machine learning, computer vision, and artificial intelligence. Also, there is growing evidence that merging image mining approaches and advanced storage technology can produce more efficient and accurate image content-based systems than traditional systems. Although the current image mining approaches are far from maturity and integrity, they open a vast space and promising research direction.

Several interesting studies involving image mining and the concepts of association rules mining have been conducted in [2], [25], [74]–[76]. On the other hand, several classification methods such as the Naïve Bayes rule, support vector machine, and decision tree have been used in the fields of data mining and machine learning [29]. Although most of them have been developed using existing approaches, the development of classification methods is still a challenging task. Besides, classification and association rules mining are two important tasks belonging to the data mining area, which can be integrated as an associative classifier to offer precise and accurate systems. Relevant works [27]–[30] demonstrate that associative classifiers can overcome some constraints introduced by other classification methods such as a decision tree that examines one feature at a time. Therefore, associative classification methods have been the focus of many studies in recent years. For example, the authors in [77] presented an associative classification method based on the Prism algorithm, which decreases the number of generated rules. In [30], an associative classification model based on a frequent fuzzy pattern has been proposed that utilizes the well-known FP-Growth algorithm. In [11], a weighted association rule-based classifier for the classification of mammograms has been proposed. The most relevant references to our research are [11], [30], [77], amongst others. Fuzzy association rules mining is gradually emerging as an appealing domain for the design and implementation of a classification framework, especially by those employing image datasets.

Fuzzy reasoning and fuzzy logic are two related but different research fields such that both are fruitful but have not yet been well linked. Probability is associated with events and not facts, and those events will either occur or not occur. There is nothing fuzzy about it. Whereas in fuzzy logic we try to capture the essential concept of vagueness. Fuzzy Logic is all about the degree of truth. Degrees of truth is used in the analysis of predicting past outcomes based on observational evidence, and are usually based on statistical beliefs that have been formed through inductive reasoning.

Specifically, Probability does not have anything to do with knowledge but helps to reason about prediction. On the other hand, prediction of future occurrence and probabilities go hand in hand and are often more complicated than first considered.

Fuzziness gives the grade or level of involvement whereas probability gives the Chance of involvement. Probability theory has nothing to reason about things that are not entirely true or false.

Fuzzy systems have received many successful applications that are not related to probability. The key reason, from our study experience, is more related to the nonlinearity formed from the fuzzy models (which like ANNs as universal approximators).

Fuzzy logic (specifically fuzzy set & rules) are useful to model human concepts and subjective understanding/knowledge and let computer improvise on it. While the probability is useful in situations with random events.

Figure 3-1 illustrates the distribution of the papers based on ARM published between 2000 and 2019 [1]. As presented, the highest number of papers were presented in 2017. This figure depicts that the evolutionary ARM has emerged as a popular topic in recent years. In the last two decades, many researchers have utilized the discovery of ARs to discover useful knowledge from the database. ARM is one of the essential components of our research. Therefore, we provide a review of ARM in terms of different aspects to supply a useful overview. We also bring up current challenges and opportunities and address potential trends and applications. One of the most important objectives of this study is to provide a reference point for researchers and data miners to be informed of state-of-the-art evolutionary ARM methods, particularly in image content. In general, the objectives of this research are as follows: (1) providing a systematic and comprehensive review of ARM; (2) developing an associative classifier for the counterfeit coin detection; and (3) specifying research gaps and proposing suggestions for directions for future research.

We indicate that ARM in the image domain has the potential to reveal significant patterns in the dataset. Investigating research applications encourages businesses, medical care systems, and governments to pay attention to the use of evolutionary methods for knowledge discovery in their respective domain.

In this research, we describe a full-fledged image mining system, named *PrFA*, that assists data miners in the complex process of extracting the units of analysis from the image database, specifying the background knowledge on the application domain and defining some form of search bias. We preserve the full power of rule mining to reduce the amount of redundant and insignificant rules by focusing on pruning methods. This aspect is particularly relevant since the number of discovered rules is usually high, and the interest of most of them does not fulfill user expectations. The new image mining tool has been applied to the coin authentication domain, thus proving the generality of the proposed solution.

3.2 *Knowledge Discovery in Databases and Data Mining*

Lately, data mining has been considered as an alternative to knowledge discovery in data (KDD), even though most studies in this field bring up data mining as a part of a whole process called KDD. Although in the last decade, there have been only a few examples of knowledge discovery in real data, nowadays, more areas profit from the exploitation of KDD techniques, such as marketing, health care systems, fraud detection applications, financial support systems, and many other domains. As a result, KDD has grown into a multidisciplinary area in computer science, including machine learning, databases, artificial neural networks, artificial intelligence, information retrieval, and data visualization.

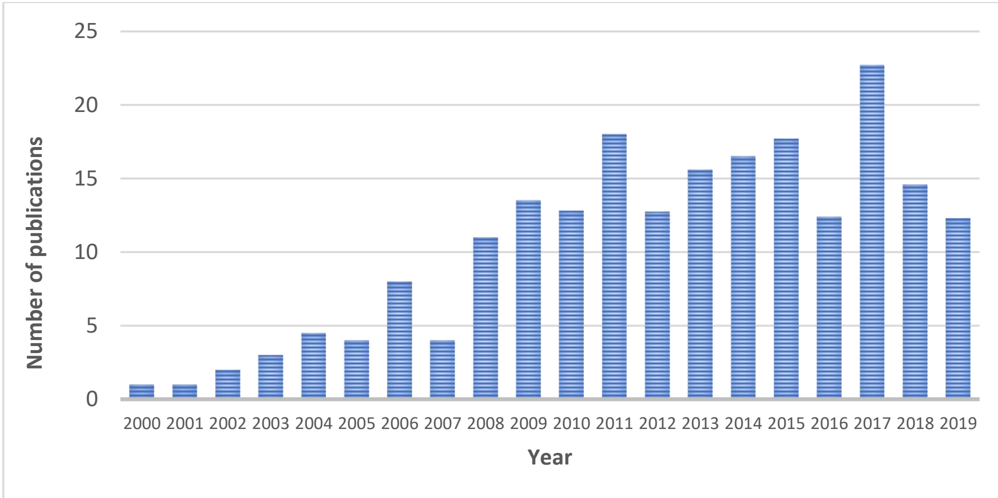


Figure 3-1 Number of papers based on ARM published from 2000 to 2019.

In general terms, KDD directs the development of tools that can automate the data analysis process. It extracts useful information and knowledge from data to help in decision making. The research in knowledge discovery in databases was proposed by Frawley et al. in 1992 [78]. They have presented a method for nontrivial extraction of implicit, previously unknown, and potentially useful information from data. According to the definition given by Todorovic in [65], four fundamental characteristics need to be available in each pattern discovered by the whole process:

- Validity: generated patterns require to be authentic for any new data with a certainty determined by interesting metrics of any methods.
- Novelty: generated patterns require to be novel compared with the previous pattern detected one.
- Usefulness: generated patterns should be significant and facilitate making decisions for the companies.
- Comprehension by users: the patterns should be understandable to the analyzer, which usually means the simplicity of knowledge.

Figure 3-2 illustrates the main KDD steps, which include data preprocessing, data mining, and post-mining [79].

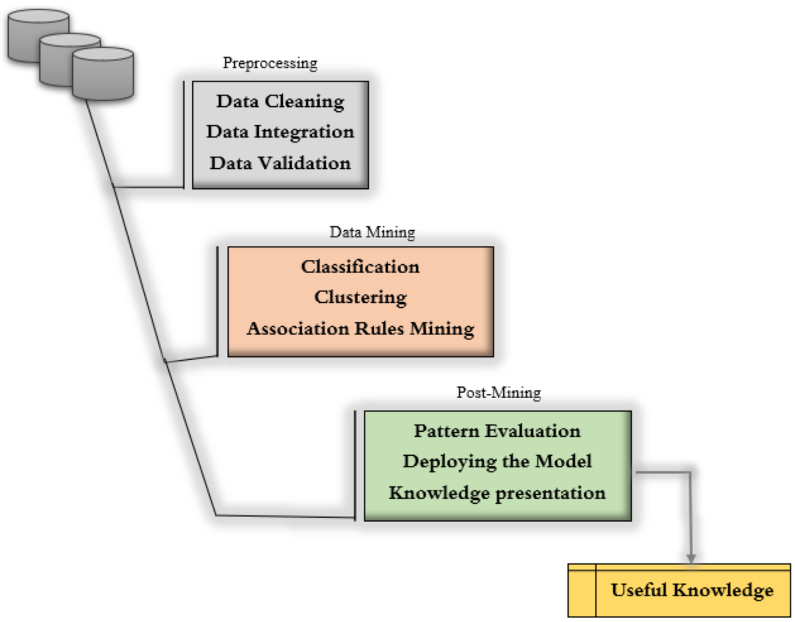


Figure 3-2 An overview of the steps comprising the KDD process.

3.2.1 *Data Preprocessing*

Data preprocessing is the first step in the whole complicated KDD process. It consists of three main tasks:

- **Data Cleaning**: In the real world, data is never 100% accurate, and it always contains missing values, noise, and additional information with errors that make data inconsistent. The data cleansing step consists of detecting, correcting or removing inaccurate and inconsistent data from the database. This phase needs considerable effort, often as much as 70% or more of the total data mining effort.
- **Data Integration**: In the real-life, data sources consist of several and with different schemas. Data integration is applied to merge resident data from different sources to provide an integrated view of data. Nowadays, this phase is considerable, especially in the commercial domain, when two or more companies share or merge their stored data.
- **Data Validation**: This phase verifies the merged data based on the two previous steps. If it has not been done correctly, a re-cleaning or re-integration will be processed.

Afterward, the data is converted into suitable forms to facilitate the application of data mining techniques.

3.2.2 *Data Mining*

Data mining is a fundamental phase in the KDD process to extract information from a dataset and transform the extracted information into a comprehensible structure for further use [1]. The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining).

To discover the hidden knowledge of the database, data mining methods are inspired by various fields, including mathematics and computer science, such as machine learning, artificial intelligence, and statistics. In the past couple of decades, data mining techniques have served many benefits for diverse areas such as business, medicine, marketing and factory assembly lines, along with others.

3.2.3 *Post Mining*

At this stage, many patterns are discovered through algorithms applied to the database with different levels of reliability. Thus, this step investigates the produced information and discover which patterns can be presented as knowledge to the user. This phase is obligatory to visualize the knowledge discovered in the data mining phase in such a form the user can interpret simply for better decision making. In general, the post-mining phase is a user-driven step in which the user becomes responsible for deciding about the interest of some knowledge. Hence, in this phase, various techniques are expanded to guide the user. For instance, the overfitting issue frequently happens with classification algorithms, in which some patterns in the training set do not exist in the dataset, and it needs user intervention to eliminate them.

3.3 *Background on Association Rules Mining*

Association rules mining is the process of finding ARs in transaction data. It is one of the most remarkable techniques in data mining, and it is considered a significant method for pattern recognition [80]. ARM is precisely the field of pattern mining aimed at extracting relationships among the features of a large database in the form of IF-THEN rules, i.e., IF some conditions are satisfied, THEN some others, too [81]. In general terms, the extracted relationships can be shown as IF-THEN statements, IF < some conditions are satisfied> THEN < some values of other attributes>. Conditions are in the IF statement called Antecedent, and those within the THEN clause are Consequences.

The number of applications that require automated data analysis methods like ARM to extract implicit, previously unknown, and potentially useful information has been increasing over the last couple of decades. Emerging data mining technology supplies different data analysis tools for a diversity of tasks, both predictive and descriptive [82].

3.3.1 *Types of Algorithms for Association Rules Mining*

There is a large number of techniques to extract association rules from databases. Although a few methods do not exactly fit in any of the categories, like [83], they are mostly cited as two main groups [84]:

- Complete Methods
- Approximation Methods (Evolutionary and swarm-based algorithms)

The first ones are algorithms, such as Apriori [85] or FP-growth [86], that can discover all rules that meet some of the requirements given. One of the important drawbacks of complete algorithms is computation requirements, and that, without an exact amount of parameter setting, these algorithms discover a huge number of rules which causes the posterior human investigation substantially effortful. As the authors in [87] have been demonstrated, extracting a huge amount of rules hides the most interesting ones.

The next groups consist of evolutionary and swarm-based algorithms, such as [88]–[90]. These groups are not expected to evoke all rules satisfying the provided criteria, but a somehow desirable sample of them. These groups resolve the drawback above of complete methods, including computation requirements and the demanding parameter setting.

Furthermore, a type of approximation method exists that has not been considered in depth for association rule mining, or at least specifically. They are heuristic operators, which, in contrast to the mentioned metaheuristics, are problem-dependent [91]. Heuristic operators are procedures that utilize certain properties of the problem, which are often computationally inexpensive to generate solutions. For instance, some heuristic operators for rule mining can be found in [92]–[95]. In this work, we propose a pruned fuzzy associative classifier based on complete methods that preserve the full power of rule mining to reduce the amount of redundant and insignificant rules by focusing on pruning methods. Our method exploits ideas for extracting and pruning rules that are expected to result in an increased interest of the experts. This framework is to consider the information gain from a powerful interpretable classification inducer and to make the rules as interpretable as possible.

3.3.2 *Definition of ARM*

The meaningful purpose of association rule mining is not to extract knowledge from data, but rather interesting knowledge. The obstacle is that interestingness is not generally well-defined in computable terms. Even more, it is commonly subjective [84].

The basic concepts of ARM are defined as follows [85]: Given an itemset I and a fuzzy transaction set T, where each transaction is a subset of I, a fuzzy association rule is said to be an "implication" of the form $A \Rightarrow B$ denoting the presence of itemsets A and B in some of the T transactions, assuming that $A, B \subset I$, $A \cap B = \emptyset$, and $A, B \neq \emptyset$. The usual metrics to establish an association rule's fitness are the *Support* (*Support*($A \Rightarrow B$), the joint probability p($A \cup B$)) and the *Confidence*

(*Confidence*($A \Rightarrow B$), the conditional probability p($B|A$)). So, a fuzzy association rule can be represented as $A \Rightarrow B$ (S, C), where S is called the *Support* and C is called the *Confidence* of rule:

$$Support\ (A \Rightarrow B) = \frac{N_{A\&B}}{N_{Database}} \qquad (3.1)$$

$$Confidence(A \Rightarrow B) = \frac{N_{A\&B}}{N_A} \qquad (3.2)$$

where $N_{Database}$ is the total number of transactions in the database; $N_{A\&B}$ demonstrates the number of transactions, which contain A and B; and N_A demonstrates the number of transactions, which contain A. Based upon the notations of *Support* and *Confidence*, fuzzy rule $A \Rightarrow B$ is an interesting fuzzy association rule, if *Support* ($A \Rightarrow B$)≥*Min_Support* and *Confidence* ($A \Rightarrow B$)≥*Min_Confidence*. The thresholds for *Min_Support* and *Min_Confidence* are set by the user [96].

3.3.3 *Types of patterns in ARM*

In this section, the basic information about different types of patterns in the ARM process will be provided as follows:

3.3.3.1 *Boolean ARM*

The definition mentioned in Section 3.3.2 is Binary ARM, which is generally recognized as BARs. In this method, frequent itemsets and rules are extracted in a Boolean dataset. Boolean ARM exclusively investigates whether an object is available or not in exchange, without considering its quality [97]. It handles only simple item-based transactions. Agrawal et al. [85] proposed the Apriori algorithm to find quickly Boolean association rules. Apriori is a well-known and widely-used ARM algorithm, which extracts ARs with high accuracy. In recent years, other algorithms like Eclat [98], have been proposed to mine Boolean association rules. In

Apriori, a breadth-first search was applied to compute the support value of itemsets, while a depth-first search using an intersection set was used in Eclat.

Although Apriori is one of the most and widely-used algorithms to mine association rules, its runtime is notably long for large datasets. Therefore, it is not capable of dealing directly with the continuous domain of data. Furthermore, Eclat and the FP-growth algorithm [86] pursue a "divide and conquer" strategy to extract rules without candidate creation [99].

3.3.3.2 *Quantitative ARM*

Extracting ARs from quantitative data is recognized as QARs. In a QAR, features can be both quantitative and categorical. Therefore, QAR is more expressive and meaningful than BARs.

In the real world, most data applications include continuous values (e.g., salary, weight, age). In contrast to the first generation of association rules, Boolean association rules, which handled only simple item-based transactions, the next generation were quantitative features. Such that, their values were elements of continuous domains includes real number domain R.

A common way to address this issue is to discretize the continuous features to different intervals. In this way, the features can be treated as categorical features. Quantitative ARM algorithms start by partitioning the feature domains and then transforming the problem into a binary one. For example, feature *A* with a value between 10 and 100 can be divided into nine intervals (10-20, 20–30, 30–40, …, 90–100). If a value is 46, the interval (40–50) becomes 1, and the other intervals remain as 0.

Although this method can solve problems introduced by quantitative attributes, it causes the "sharp boundary" problem. The sharp boundary problem either ignores or over-emphasizes the elements near the boundary of intervals in the mining process. Hence, the major problem with discretization is the loss of information and negligible results [100]. Besides, the efficiency depends on the distinctive intervals, while specifying fit intervals is solid.

3.3.3.3 *Fuzzy ARM*

The sharp boundary between intervals is a major obstacle with discretization-based QRs, which leads to loss of information. In this section, we intend to demonstrate the sharp boundary problem using an example. Assume that there are three intervals for the "Salary" feature, "LOW" [0, 30K], "MEDIUM" [30K, 60K], and "HIGH" [60K, 90K]. Consider a person with a 29.9 K income. According to the Quantitative description, this person will be classified into the "LOW" interval. Nevertheless, it is uncaused to classify a person with a 29.9 K income into the "LOW" group when an individual with a 30.1K income is classified into the "MEDIUM" group. This issue is referred to as sharp boundaries.

As a remedy to the sharp boundary problem, the fuzzy set concept, introduced by Zadeh [101], has been used more frequently in mining quantitative association rules. The fuzzy set theory is the utilization of natural languages to indicate the concepts. This technique can create linguistic knowledge because it converts quantitative values into linguistic terms [102]. Since the fuzzy set theory resembles human reasoning, it is an intelligible and admired technique in intelligent systems. The application of fuzzy concepts is terrific for expert systems that aim to exploit a human-like way in making-decision and programming. This approach is better than the partitioning method because fuzzy sets provide a smooth transition between members and non-members of a set and increase the flexibility of systems.

On the other hand, In the objective world, the information is mostly vague and indefinite, so the research on the fuzzy concept has important practical significance [103]. In the information society, the data collected in many real systems are inaccurate. If this uncertain information cannot be properly analyzed, it will probably lead to a great error between the inference result and the objective fact.

The important advantages of applying fuzzy set concepts are as follows: (a) Capacity to represent inherent uncertainties of human knowledge with linguistic variables; (b) Smooth interpretation of the results, because of the representation of the natural rules; and (c) Easy development of knowledge through the enhancement of new rules.

In general terms, fuzzy ARM is the process of applying the fuzzy set and Membership Functions (MFs) concepts in mining QARs. Since fuzzy ARM performs a smooth transition between a member and a non-member of a set, it produces the result more interpretative than crisp ARM [104], [105]. Unlike the crisp sets in which each element represents its presence and absence in an interval, fuzzy sets determine the degree of a value belonging to an interval. By applying the fuzzy sets theory for the feature of 'Salary', we synthesize the three intervals with MFs by developing the Boolean values 0 and 1 to the continuous values from 0 to 1 ([0, 1]). The values of 1 and 0 illustrate presence and absence, respectively [106]. Figure 3-3 displays an example of crisp sets (Figure 3-3 (a)) and fuzzy sets of the 'Salary' feature (Figure 3-3 (b)). According to Figure 3-3 (b), when a person's salary is 25K, it has a membership of 0.76 in the fuzzy set "MEDIUM" and a membership of 0.29 in the fuzzy set "LOW".

In FARs, numeric feature values are demonstrated in terms of linguistic terms. MFs are first applied to convert each numeric feature into linguistic terms. Then, the scalar cardinality of each linguistic term is computed on all transactions. Finally, the fuzzy counts-based mining process is executed to discover Fuzzy ARs (FARs). In FARs, each feature x_i is demonstrated at least by two fuzzy sets, such that there is one membership function per fuzzy set. To specify the degree of membership of each value of x_i, the membership function is assessed.

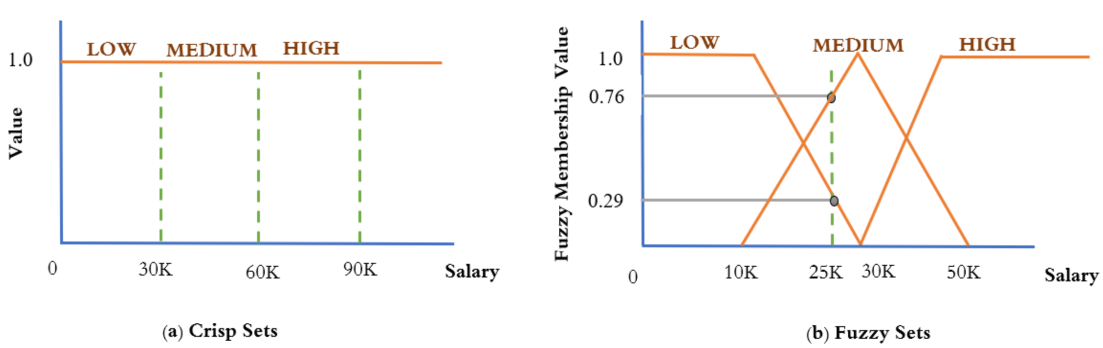


Figure 3-3 An example of crisp sets (a), and fuzzy sets (b) for the 'Salary' feature.

In this subsection, we briefly present a supreme view on the fuzzy set theory. According to the definition in mathematics, the elements in fuzzy sets have membership values. Fuzzy sets theory is now widely used in different fields, such as linguistics, decision systems, and cluster analysis. The fuzzy sets theory uses a membership function to describe the fuzzy relation. The range of the membership function value is the unit interval [0,1] [103].

Fuzzy sets can be represented as (U, m), wherein the domain U is a set, and m: $U \rightarrow [0,1]$ is a membership function. For each $x \in U$, the value of $m(x)$ is called the membership degree of x in (U, m). for the finite set $U = \{x_1, \ldots, x_n\}$, the fuzzy set (U, m) is usually represented by $\{m(x_1)/x_1, \ldots, m(x_n)/x_n\}$. For a given domain U, a mapping m_A: $U \rightarrow [0,1]$ can determine a fuzzy subset A, whose representation is as follows:

$$A = \begin{cases} \sum_{x_i \in U} \frac{m_A(x_i)}{x_i}, & U \text{ is a finite set } \{x_1, \ldots, x_n\} \\ \int \frac{m_A(x)}{x}, & U \text{ is an infinite set, and } x \in U \end{cases} \tag{3.3}$$

where m_A is the membership function, and $m_A(x)$ indicates the membership degree of x to A. Note that $\sum_{x_i \in U} \frac{m_A(x_i)}{x_i}$ is not an expression of the sum of fractions, but merely a sign [103].

Furthermore, a definition of Fuzzy ARs is defined as follows [92]: If $X=\{x_1, x_2, \ldots, x_p\}$ is $A=\{f_1, x= f_2, \ldots, f_p\}$ then $Y=\{y_1, y_2, \ldots, y_p\}$ is $B=\{g_1, g_2, \ldots, g_p\}$, where A and B consist of the fuzzy sets related to the features in X and Y, respectively. f_i *and* g_j are the fuzzy sets associated with features x_i and y_j, respectively. The rule antecedent is "X is A", while the rule consequent is "Y is B". The fuzzy support value of itemset Z and its fuzzy sets F, denoted as $S_{<Z,F>}$, is computed as shown below:

$$S_{<z,F>} = \frac{\sum_{t_i \in T} \prod_{Z_j \in Z} \mu_{z_j}(f_i \in F, t_i[Z_j])}{|T|} \tag{3.4}$$

where $|T|$ is the number of transactions in the dataset.

In recent years, many researchers have made a thorough study of fuzzy set concepts. Despite the wide use of association rules mining, the research on fuzzy-based association rules mining is relatively few. It is worthy to note that the fuzzy concept is an excellent tool to extract association rules. In this research, the use of fuzzy association rules is considered as the key component of the proposed approach because of the affinity with the human knowledge representation.

3.3.3.4 *Class ARM*

ARM and classification are two widely-used techniques in the real world [107]. The integration of these models is a new trend in data mining. In general, classification and association rules mining are two important tasks belonging to the data mining area, which can be integrated as an associative classifier to offer precise and accurate systems. ARs have been successfully utilized for classification, and the simulation results demonstrated that it could achieve undertaking and promising accuracy. Relevant works indicate that since ARM discovers highly confident associations among multiple features, associative classifiers can overcome some constraints introduced by other classification methods such as a decision tree that examines one feature at a time. Unlike an association rule that contains different items in the consequent part, a classification rule includes the class feature only [108]. A general strategy to discover class ARs is to extract ARs in the first stage, and then ARs with only the class feature in the rule consequent is specified as class ARs. These patterns are efficacious in domains that require to prediction or classify a certain target. Therefore, associative classification methods have been the focus of many studies in recent years. For instance, class ARs are applied in medicine to predict the chances of breast cancer occurrence based on a given dataset [109]. The authors in [77], presented an associative classification method based on the Prism algorithm, which decreases the number of generated rules. In [30], an associative classification model based on a frequent fuzzy pattern has been proposed that utilizes the well-known FP-Growth algorithm. In [11], a weighted association rule-based classifier for the classification of mammograms has been proposed. The most outstanding references for the associative classifier found in the literature are [11], [30], [77], amongst others.

Fuzzy association rules mining is gradually emerging as an appealing domain for the design and implementation of a classification framework, especially by those employing image datasets. In this research, we design a pruned fuzzy associative classifier that has an optimization procedure of search for image features and strong fuzzy rules that better satisfy the desired measures for counterfeit coin detector system.

3.4 *Image Mining*

Recent years have witnessed tremendous growth in image databases, and the World Wide Web has emerged as a spacious repository of image data in the world [20]. However, owing to a large number of images in these databases, the only viable means to knowledge discovery and image classification is to automate this process, since the manual task can be a tedious and expensive job. Impressive product design is a challenge in the industry, where companies need to expand high-quality products that convince their customers. This procedure not only elevates customer satisfaction but leads to the achievement of companies that utilize this competitive edge. ARM is currently one of the most active data mining topics. Over the last few years, it has achieved enormous success in a plethora of decision-making systems in the industry. Research studies have derived the advantage of ARM to recognize patterns in various fields. According to the investigation proposed in [1], most of these publications are in the fields of security in computer networks (20%), healthcare (17%), and industry (14%). Figure 3-4 demonstrates the distribution of the algorithms in different fields. This figure shows the importance of data mining techniques, especially ARM.

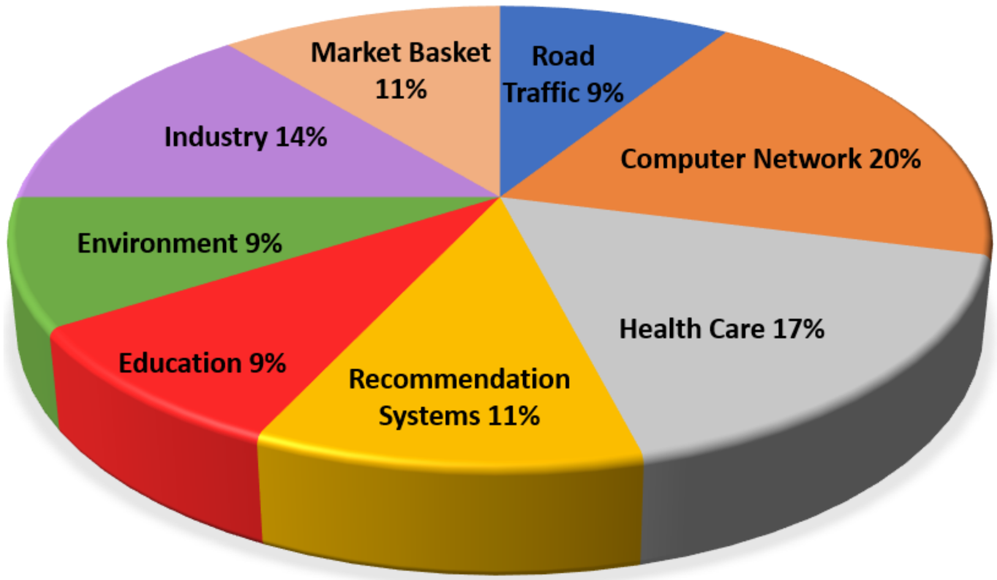


Figure 3-4 Distribution of ARM in different domains [1].

ARM for images is considered a relatively nascent subfield. The focus of image mining is concerned with the extraction of valuable information that is not implicitly accumulated in the images. Generally, there are two principal approaches to extract association rules based on image content. The first one includes merely mining images for feature extraction, while the second one involves mining images along with finding information associated with the image's category. In this research, the latter approach will be applied. We develop an image mining system on top of the concept of fuzzy association rules using a blob detection method and relationship predicate that helps us to discover the implicit information from the images in a way closer to the human's viewpoint.

It is said that an image is worth more than one million pixels, but specifying the likely groups of pixels that form the substantial information of the image is considered to be a challenging problem by the image mining process. Knowledge derived from the image mining domain can be more descriptive compared to other areas. Conventional approaches mostly use image features and develop feature-based clusters. Based on this fact, we conjecture that fuzzy association rules mining can capture more information from the image dataset. Hence, in this research, we present a method of formulating this conjecture as a general form to provide a common framework for image classification appearing in other domains.

3.5 *Summary*

In this chapter, a general background on Knowledge discovery in the database process has been described. The main focus of this chapter has been conducted based on the preliminary concepts and notations of the field of association rule mining. Furthermore, the fundamentals information about the various types of patterns in the ARM process includes Boolean, Quantitative, Fuzzy, and Class ARM has also been provided.

Chapter 4

Proposed Framework

This chapter illustrates the proposed framework for counterfeit coin detection. It is organized as follows. An overview of image mining is presented in Sect. 4-1. In Sect. 4-2, a schematic view of our proposed framework, *PrFA*, is introduced in detail. In Sect. 4-3, the processing module is presented. In Sec. 4-4 the image miner module is discussed. In Sect. 4-5, post-processing of two pruning techniques based on redundancy restriction and feature dimensionality reduction are presented. In this section, we propose an engine based on a fuzzy associative feature selection method named Engine_*FAFS,* which is incorporated into the proposed framework to control the size of the feature subset of images. In Sect. 4-6, our proposed fuzzy associative classifier is introduced in detail. In Sect.4-7, the optimization procedure with a summary of the computations for the *PrFA* framework, which includes the application of the PSO algorithm is discussed. Finally, this chapter is concluded in Sect. 4-8.

4.1 *Overview*

The focus of image mining in the proposed method is concerned with the extraction of valuable information that is not implicitly accumulated in the images. It is worthy to note that fuzzy association rules mining based on image content is feasible and gives strong rules that can be further used for the effective classification of coin images. Indeed, the primary concern of the proposed framework is to perform the top-down deepening search progressively and to extract relationships among the features of the coin images. Our proposed framework exploits the correlation between low-level features in an image and high-level semantic content to classify the images. Besides, it shows that adequate similarities exist to provide significant fuzzy rules, which can then be exploited for auto-classifying of coin images.

We design a pruned fuzzy associative classifier that has an optimization procedure. It searches for image features and strong fuzzy rules to satisfy the desired measures for counterfeit coin detector system. Our pruning advantage over regular associative classifier methods is that it avoids the overfitting for the rules mining algorithm. To obtain an accurate fuzzy associative classifier, we apply two post-pruning techniques based on redundancy restriction and feature dimensionality reduction. In this research, an engine based on a fuzzy association feature selection method will be proposed, which is incorporated into the proposed framework to control the size of the feature subset of images. Because this is the motivation behind our project, we termed the framework *PrFA*: a framework for image-based counterfeit coin detection using Pruned Fuzzy Associative Classifier.

4.2 *PrFA Architecture*

In the proposed framework, we extract fuzzy association rules, which demonstrate the frequent pattern that occurs together in similar types of coin images. Our proposed framework for counterfeit coin detection is composed of two main modules; (*Module*#1: Processing) and (*Module*#2: Image Miner).

The preprocessing is necessary to modify the quality of coin images and make the feature extraction method more reliable. The original coin images should be transformed into a suitable format to extract fuzzy association rules. In the preprocessing step, the median filtering is used to remove digitization noise. To extract feature descriptors, especially for poor-quality or damaged coin images, we need to produce an image with brighter regions. To address this issue, we calculate the new image by adding the original image to a coefficient of its gradient magnitude, and then we extract feature descriptors in the new image.

In the first module, the digital coin images are segmented to find the region of interest (ROI). An important aspect of the first module is that it recognizes the nature of images as combinations of the set of localized coherent regions based on visual similarity and allows us to access the images at the level of blobs. By extracting image regions that correspond to the blobs, the features are extracted and organized as inputs for the next module. In the second module, the frequent patterns in the images using fuzzy association rules mining are extracted.

PrFA firstly applies the blob detection method to extract the features of images. Then, it mines fuzzy association rules, which facilitates their interpretation in linguistic terms and avoids unnatural boundaries in the partitioning of the domains of the coin image features. Afterwards, the rules are prescreened, selected and tuned with the evolutionary post-processing pruning algorithms to get a compact, efficient set as the final classifier. In contrast to other approaches, which work with traditional feature extraction methods and thus require to discretize the database, *PrFA* utilizes the blobs detection method by generating and tuning its fuzzy linguistic terms.

The fuzzy association rules can discover relationships among patterns in image datasets. So, each input digital coin image is associated with a keyword, i.e., genuine or fake. The system is performed on descriptors created from the image content that contains information about colour, texture, shape, and size. The similarity of these attributes determines the blobs. Each detected blob is assigned an ID, and fuzzy association rules are generated based on the presence/absence of blobs in images. The extracted blobs are considered as items to generate fuzzy association rules, and images are considered as transactions. Extracted rules represent implicit knowledge contained in images and implicit relationships that exist in a set of images. This knowledge can assist in classifying images. In this research, we propose a fuzzy associative classifier to classify the coin images. The process of classification consists of four stages: Generating all Fuzzy rules, Extraction of Pruned Rules, Building Classifier, and Predicting Class. Our fuzzy associative classifier is an emerging zone that unifies supervised classification techniques and fuzzy association rules mining. While the objective of association rules mining is to extract interesting relationships among features in datasets and the aim of classification problems is the prediction of class labels, their fusion leads to an interdisciplinary endeavour that draws upon an accurate associative classifier. Figure 4-1 shows a schematic view of the proposed framework. The details of this framework are described in the following subsections.

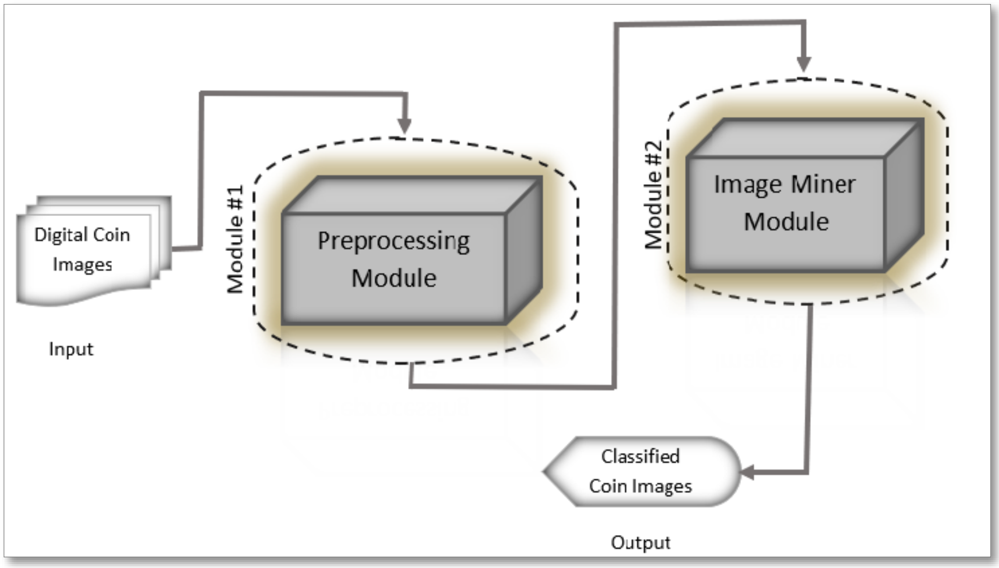


Figure 4-1 Schematic view of the proposed framework.

4.3 *Processing Module* (*Module* #1)

In the processing step, the goal is to obtain and preprocess information about pixels in a given neighbourhood that are related to one another. We aim to bridge the gap from the low-level image properties to the objects, which can ultimately be considered as an object recognition problem. In this step, a method will be applied for image retrieval based on segmentation into regions and querying using properties of these regions. Feature extraction is the most significant component of designing an intelligent system based on image pattern recognition since even the best classifier will run inaccurately if the features are not chosen well.

The main part of image mining is to identify similar objects in different images. The segmentation of images to find ROIs is one of the most important processing steps in image mining. By extracting ROIs that correspond to the blobs, we access the images at the level of objects rather than global image properties. In this framework, we intend to capitalize on the detection of blobs to find objects in images automatically. Figure 4-2 indicates samples of genuine and fake coins of years, 1990, 1991, 1996, and 2008.

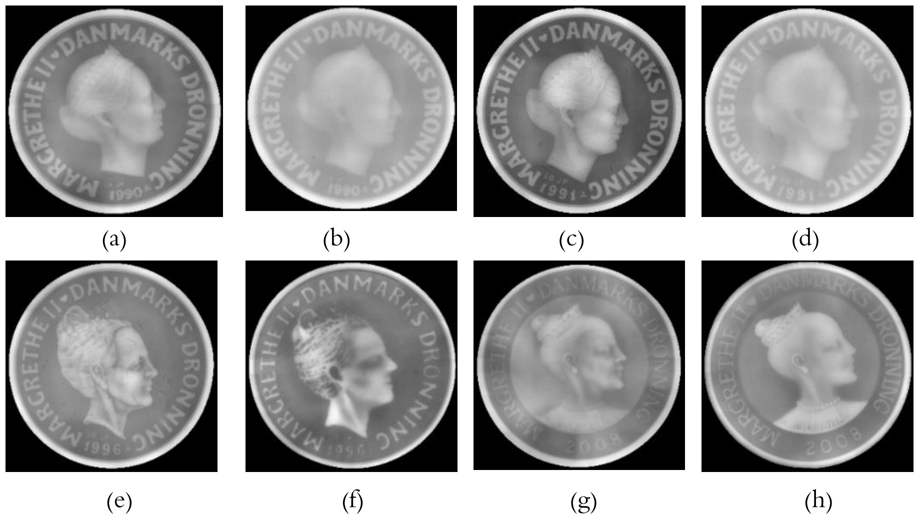


Figure 4-2 Examples of genuine and fake coins: (a) genuine 1990, (b) fake 1990, (c) genuine 1991, (d) fake 1991, (e) genuine 1996, (f) fake 1996, (g) genuine 2008, and (h) fake 2008.

In this framework, a method to detect ROIs is incorporated into the proposed framework that focuses on blob detection. The approach to blob detection is structured around a sequence of increasingly specialized grouping activities that produce a "Blobworld" representation of an image, which is a transformation from the raw pixel data to a small set of localized coherent regions in colour and textual space. "Blobworld" technique has been remarked as a noticeable method for object recognition that is based on image segmentation using the Expectation-Maximization algorithm on combined colour and texture features [72].

In the processing module, the pixels will be grouped into regions by modelling the joint distribution of colour, texture, and position features. In this manner, the EM[1] algorithm is applied to estimate the parameters of a mixture of Gaussians model of the joint distribution of pixel colour

[1] Expectation-Maximization

and texture features. This approach is related to the MDL[1] principle to perform segmentation based on motion. Our coin image dataset should be transformed into a suitable format to extract fuzzy association rules. Therefore, we first use a blob detection method, and then we combine these blobs with their descriptors to build a new data set for the rule mining process. The details needed to perform *Module#*1 are illustrated in Figure 4-3.

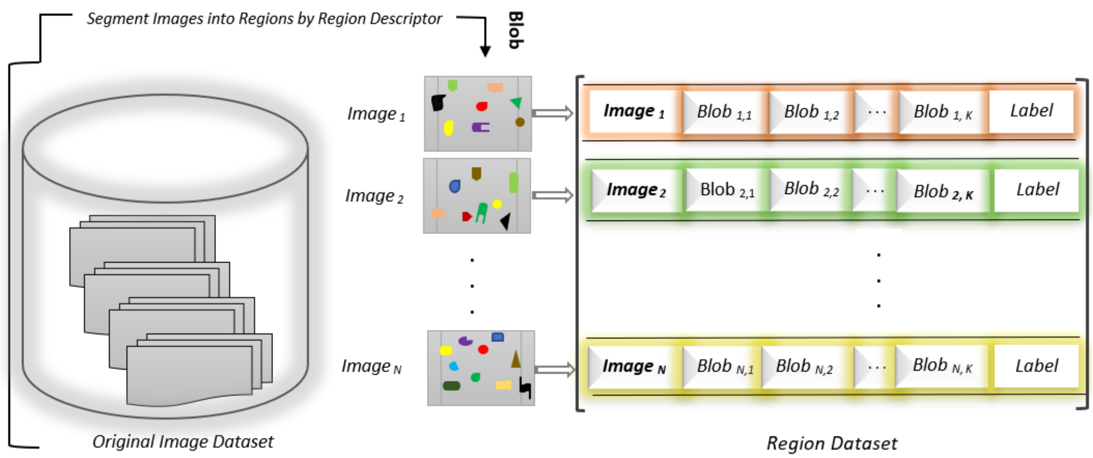


Figure 4-3 Details of Processing (*Module#1*).

Based on the type of analysis carried out, blob detection is the most suitable method for our approach. Since, some other feature extraction methods that use global properties, generally correspond to objects or parts of objects. These methods identify specific objects drawn from a finite collection, but they are not usually effective at the general image analysis task, which requires both image segmentation and image classification.

Previous methods for feature extraction such as edge-based features are extremely dependent on the type of coins. Therefore, providing a method to analyze the edge-based features can be useful for only one type of coin. For example, the distance between two letters on a coin, the length or width of a letter or a shape on the coin etc. are different in various coins. Figures 4-3 (a) and (b) show two genuine Danish 1991 coins with poor and good quality respectively and Figures 4-3 (c) and (d) show the edges detected by canny edge detector with the same parameters. Apart from

[1] Minimum Description Length

this important objection, as can be seen in Figure 4-3, the edges on the coins are worn out or abrasive and cannot be useful even if we build our system for only one type of coin.

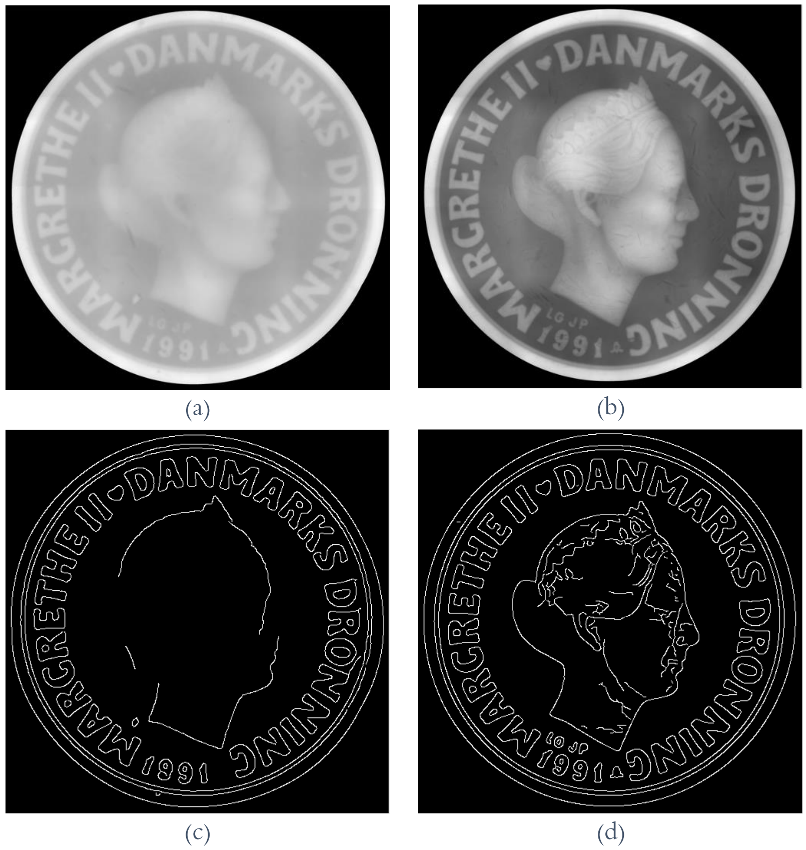


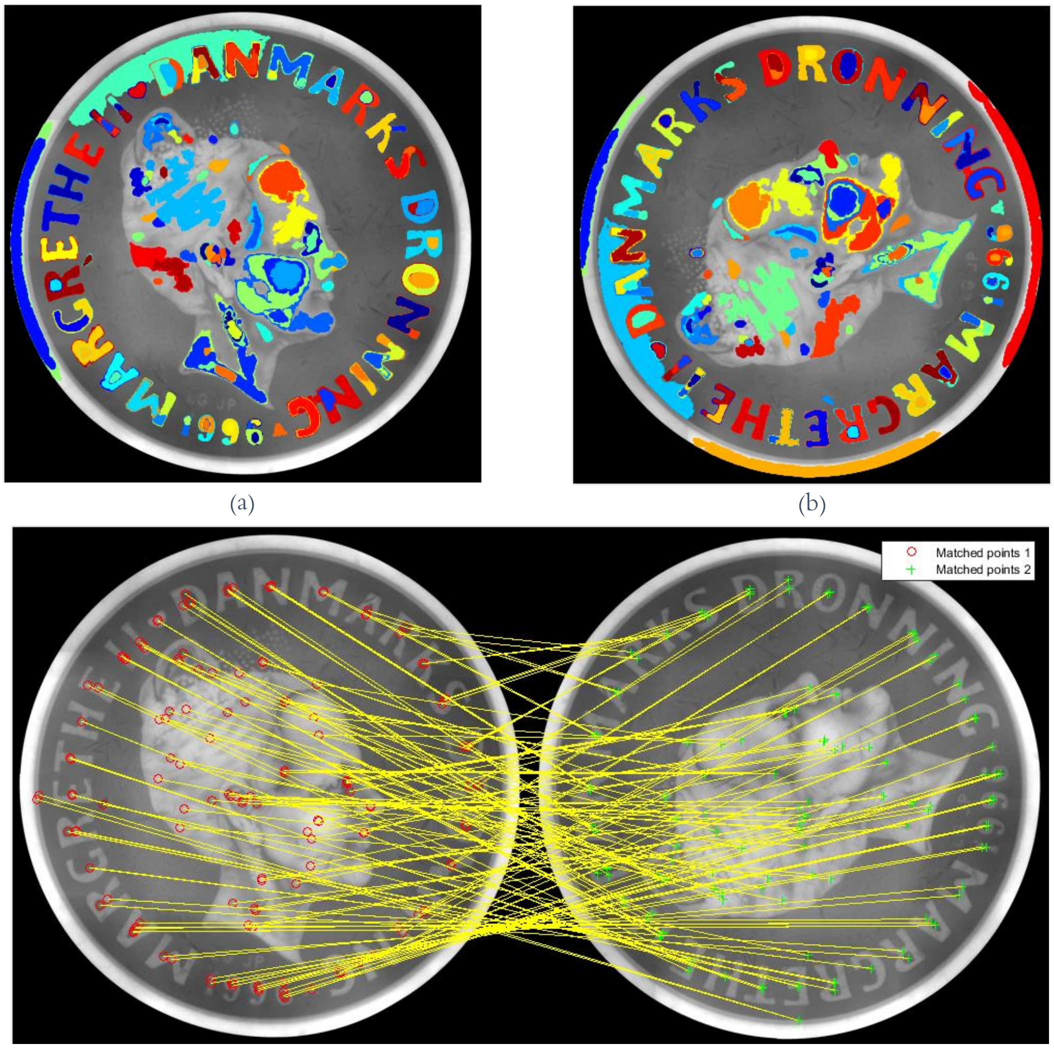
Figure 4-4 Edge detection for two coins with different qualities; (a) a poor quality Danish coin, (b) a good quality Danish coin, (c) and (d) edges detected by Canny for a and b, respectively.

From the different methods utilized for feature extraction, we have selected the blob detection method because it recognizes the nature of images as a combination of objects, and querying is more meaningful than it is with simple representations. In simple words, it bridges the gap from the low-level coin image properties ("stuff") to the high-level properties ("things").

Among the advantages of the blob detection method, we highlight two major ones related to our research:

1. One main reason to apply the blob detection method is that it provides information about regions of interest for further processing, which cannot be simply obtained from edge detection or corner detection methods. In our research, the blob detection assists the framework to recognize the nature of images as combinations of the set of localized coherent regions based on visual similarity. Therefore, it allows us to access the images (Genuine and Fake) at the level of objects rather than global image properties. In this method, the pixels can only be linked when their features are stable over scales. Such processes could lead to promising and reliable image segmentation. In this research, we can accurately locate the blob centers but can also estimate the scales, shapes, and orientations of the detected blobs.
2. Another reason to use a feature extraction method based on blob detection is that the feature extraction algorithm is rotation invariant because of the nature of blob detection. Since scanning a coin is manual, the coins are not registered well. Therefore, the rotation process is one of the most important preprocessing methods that must be performed before any feature extraction from the coin surface. Figure 4-5 Illustrates the blobs for an image and its rotated counterpart can be detected without the need of rotation while extracting features using edge detection is perfectly dependent on the position of the edge on the images.

Since the IBIS TRAX-3D scanner that we applied for image inquisition could precisely digitize both 2-D and 3-D images, we capitalize to detect the blobs by considering the height or depth information to test our model, as well. More information about the different results for blob detection methods will be presented in Chapter 5. The blob detection is one of the principal parts of the process. Therefore, we accomplish an evaluation of three different techniques for finding blobs. We also compare the performance of the proposed classifier by three different state-of-the-art blob detection algorithms, including Blobword, MSER, and watershed (Chapter 5). This comparison helps us to find the best blob detection method, which is compatible with our framework.



(a) (b)

(c)

Figure 4-5 Rotation invariance of feature based on a blob detection algorithm. (a) A set of blobs detected on a 1996 Danish coin, and (b) A set of blobs detected on a 1996 Danish coin rotated by 90 degrees.

In this research, we illustrate that our processing module enables the extraction of image features in such a way that the output can be used for rules mining. Firstly, all pixels will be subdivided into regions by modelling the joint distribution of texture, colour, and position features with a mixture of Gaussians using Expectation-Maximization and the Minimum Description Length principle. Secondly, the regions will be defined based on texture and colour properties. Finally, these regions can be accessed as an item for association rules mining.

Figure 4-6 illustrates these steps for a sample image from pixels to feature extraction. It is necessary to mention that the python module called *skimage* was used for blob detection in this research. This module is based on detecting the Laplacian of the Gaussian for candidate areas in the image. Laplacian of the Gaussian finds a gradient from a given central location and tries to determine an edge of the blob based on differences in the gradient. Therefore, each image will be converted to greyscale and then displayed as an array. This array will then be forwarded as input to the Laplacian of Gaussian function in *skimage*. The processing result of this study is a list of (x, y, r) coordinates, where (x, y) shows the center of the blob and r indicates the radius of the blob. After blob detection, we create a new dataset, such that each transaction represents one image with the visual blobs extracted along with the class label. The new transactional dataset is submitted to the image miner module as input for the rules mining step. Figure 4-7 demonstrates a sample of the processing step in finding the blobs of an image. The colourful regions illustrate the extracted blobs. In this research, we exploit the diverse colours merely for visualization. We model the images with a set of transactions, such that each transaction represents one image with the visual blobs extracted along with the class label. In this way, the original image dataset will be converted to the transactional dataset, where the first value in each row is an image id, and the second value onward is their blob's information. So, each image formally is represented as a vector of blobs. It should be noted that N represents the number of original images. Blobs in one image are compared based on similarity function to blobs in any other images, and then each blob is labelled with a cluster to form a transactional dataset.

The labels of the input coin's image, i.e., G(Genuine) or F(Fake), and the extracted blobs are used to build the transactional dataset. Indeed, the features extracted are organized in a dataset, which is the input for *Module*#2 of the proposed framework.

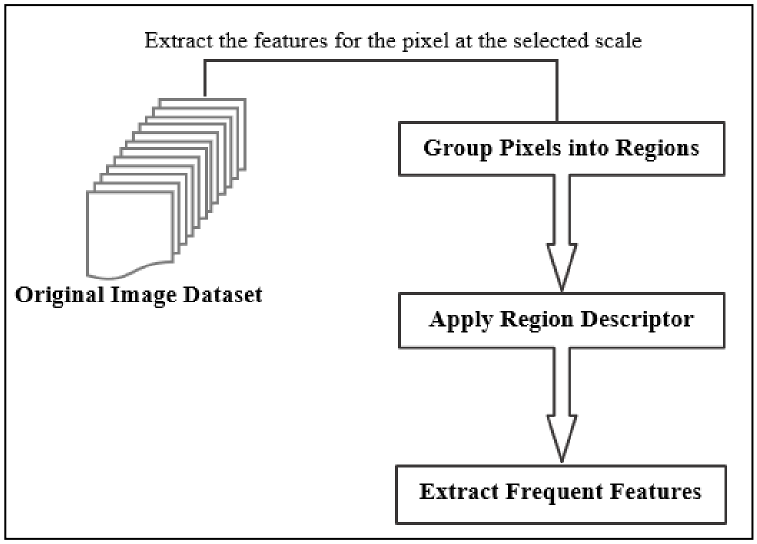
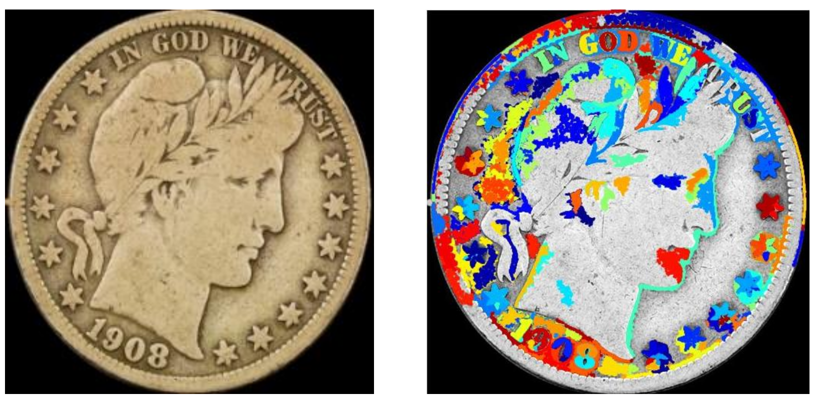


Figure 4-6 Processing steps for a sample image from pixels to feature extraction.



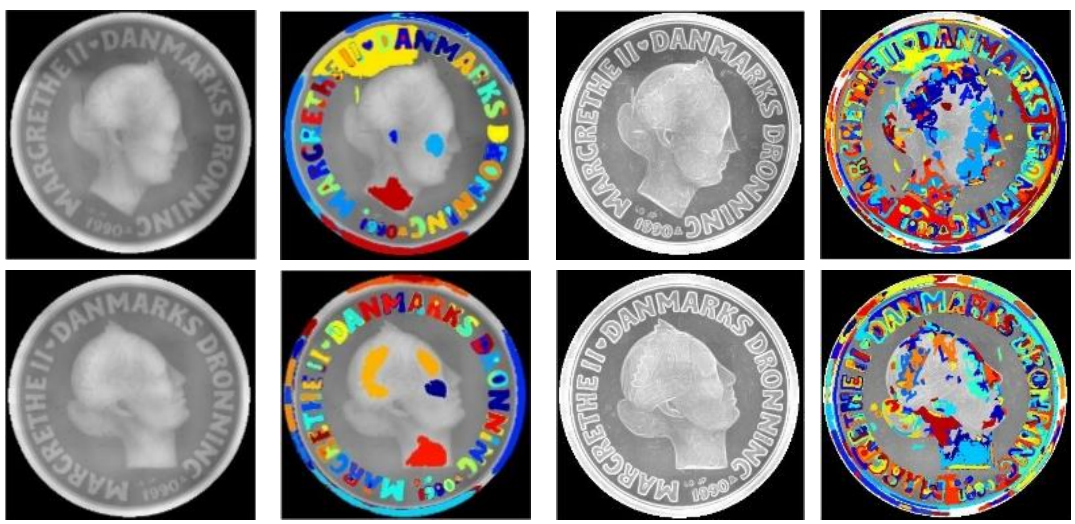
(a) (b)

Figure 4-7 Processing phase on an example image (one Dollar US coin): (a) original image; (b) preprocessed image after blob detection.

As mentioned earlier, in the preprocessing step, the median filtering is used to remove digitization noise. To extract feature descriptors, especially for poor-quality or damaged coin images, we need to produce an image with brighter regions. To address this issue, we calculate the new image by adding the original image to a coefficient of its gradient magnitude, and then we extract feature descriptors in the new image. Figure 4-8 shows two different results in finding interest regions of an original and preprocessed image. After finding the blobs in each image of the image dataset, the next step will be employed to select the regions of interest of each image, which are seen in at least one of the other images. It means that each blob of an image would remain in the *blob-set* of the image if there is a match in the pairwise comparison of all images' *blob-set*.

Figure 4-9 illustrates the matched pairs of two different Danish 20 Kroner 1990-coin images. The blobs which are not paired in any comparison will be removed from the *blob-set*. Therefore, we have the final blob dataset, also called the *filtered blob-set*. This process is the first filtering process in the proposed method. The second filtering process, which also leads to labelling the blobs, uses the fuzzy c-means clustering algorithm [110]. In this process, we extract feature vectors of blobs for creating the *final blob-set* and apply fuzzy c-means on them. We also calculate several attributes of blobs, such as *location*, *axes*, *orientation*, *area*, the *median of the colours*, and the *mean of the colours,* which are used as discriminative in the fuzzy clustering algorithm.

Figure 4-10 demonstrates blob detection without filtering, with the first filtering process, and the last filtering process, respectively.



(a) (b) (c) (d)

Figure 4-8 Effect of the filtering process on the blob detection: (a) original images, (b) blob detection without preprocessing, (c) original images after preprocessing, and (d) blob detection on pre-processed images.

Figure 4-9 The matched pairs of two different Danish 20 Kroner 1990-coin images.

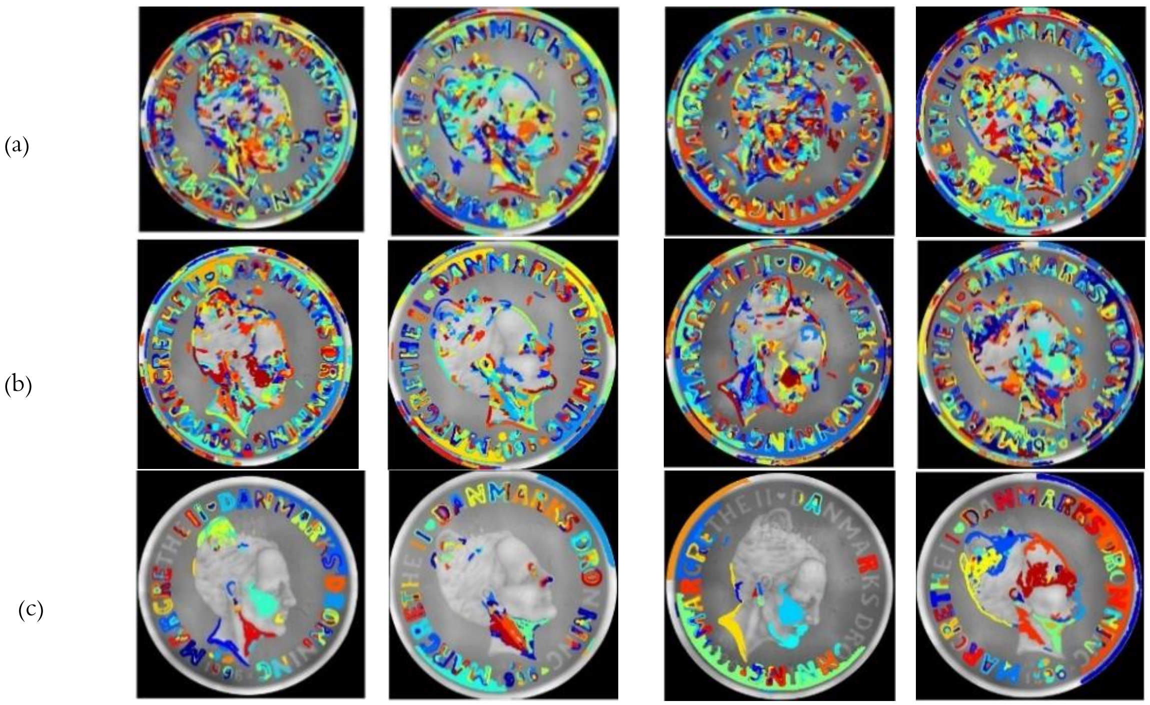


Figure 4-10 Blob detection: (a) blobs before removing unwanted regions, (b) removing unwanted blobs after feature matching, and (c) removing unwanted blobs after fuzzy clustering.

Suppose that we want to categorize the blobs of different images into n clusters. First, we filter out the blobs if the maximum membership degree of the blob $B_{i,j}$ is smaller than a threshold ($\max\left(\mu_1^{B_{i,j}}, \mu_2^{B_{i,j}}, \dots, \mu_n^{B_{i,j}}\right) < \tau$); where $B_{i,j}$ is the jth blob of an image i and $\mu_k^{B_{i,j}}$ is the membership degree of the blob to the kth cluster. Therefore, the blob $B_{i,j}$ is in cluster k if $\mu_k^{B_{i,j}} = \max\left(\mu_1^{B_{i,j}}, \mu_2^{B_{i,j}}, \dots, \mu_n^{B_{i,j}}\right)$ and $\mu_k^{B_{i,j}} > \tau$. τ is a threshold that is determined by the user. In this research, we apply the Particle Swarm Optimization (PSO) algorithm to characterize the optimum value of τ. Details about these experiments will be provided in the experimental setup and results chapter.

Regarding Table 4-1, supposing $\tau = 0.5$, the first blob of image number one and the second blob of image number three are removed from the blob-sets of the images. After the fuzzy clustering process, we also label each remaining blob by the number of the cluster which belongs to, intending to create the itemset and the fuzzy transaction dataset for the fuzzy association rule mining step. Table 4-2 shows an example of the fuzzy transaction dataset whose values are the number of repetitions of the blobs in the images.

Table 4-1 The result of the fuzzy clustering algorithm for 3 clusters and a few unlabeled blobs.

	Membership Degree of Blobs								
Number	$B_{1,1}$	$B_{1,2}$	$B_{2,1}$	$B_{2,2}$	$B_{3,1}$	$B_{3,2}$	$B_{3,3}$	$B_{3,4}$	$B_{5,1}$
1	0.3	0.6	0.05	0.4	0.03	0.23	0.85	0.001	0.3
2	0.32	0.12	0.91	0.58	0.1	0.29	0.0	0.009	0.66
3	0.38	0.28	0.04	0.02	0.87	0.48	0.1	0.9	0.04

Table 4-2 An example of the fuzzy transaction dataset.

Images	**Number of Objects**									
	$O_{1\ low}$	$O_{1\ medium}$	$O_{1\ high}$	$O_{2\ low}$	$O_{2\ medium}$	$O_{2\ high}$	$O_{3\ low}$	$O_{3\ medium}$	$O_{3\ high}$	**Class**
$Image_1$	2	0	1	0	0	3	1	1	1	Fake
$Image_2$	1	2	0	2	3	1	1	2	1	Genuine
$Image_3$	3	1	2	1	0	2	1	0	1	Fake
$Image_4$	1	1	0	3	3	1	1	1	0	Genuine

4.4 *Image Miner Module* (*Module* #2)

The image miner module is the most crucial component of the proposed system. This module aims to discover implicit information from the coin images. The details needed to perform *Module*#2 are illustrated in Figure 4-11. At this point, the images are considered as a transaction of features. A fuzzy association rule in coin images is a rule that associates features and the relationship among them in images. In this research, the *FGBRMA* algorithm is used to mine fuzzy association rules [111]. These rules will be used to build a fuzzy associative classifier, which is then applied to classify the digital coin images into two categories: genuine and fake. The *FGBRMA* is an effective algorithm since it scans the dataset only once and applies Boolean operations on tables to generate fuzzy grids and fuzzy association rules. The division of the features into many fuzzy partitions is widely used in pattern recognition problems. It is noted that fuzzy sets provide a smooth transition between members and non-members of a set and increase the flexibility of systems.

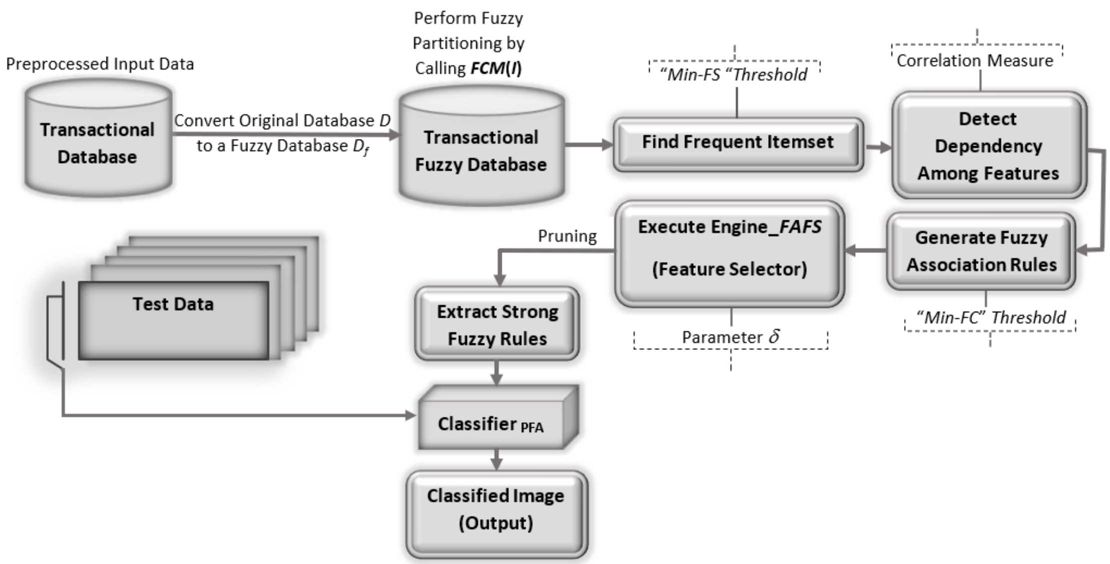


Figure 4-11 Details of Image Miner (*Module#2*).

4.4.1 Fuzzy Association Rules Induction

For our methodology, we use fuzzy grids–based rule mining algorithm (FGBRMA) is used to mine fuzzy association rules [111]. This algorithm includes two main steps: large fuzzy grids generation and fuzzy association rules generation. In the first step, this algorithm applies the fuzzy partition method for each feature so that both quantitative and categorical features are divided into *K* different linguistic values (*K*=2, 3, …). Each feature is considered as a linguistic variable, and the variables are divided into different linguistic values. Each linguistic value can be applied to demonstrate a candidate 1-dimensional fuzzy grid. Also, it can use two large 1-dimensional fuzzy grids to create a candidate 2-dimensional fuzzy grid to produce a candidate high-dimensional fuzzy grid. The fuzzy *Support* (called *FS*) is computed (Eq. (3.1)) to check whether this fuzzy grid is large or not. When its *FS* is larger than or equal to the predetermined Minimum Fuzzy *Support* (called *Min_FS*), it can be said that it is a large *k*-dimensional fuzzy grid. When all of the large fuzzy grids have been discovered, the next step will start. In the second step, each fuzzy rule *R* is created by two large fuzzy grids. The fuzzy *Confidence* (called *FC*) is computed (Eq. (3.1)) to check whether this rule is effective or not. When its *FC* is larger than or equal to the predetermined Minimum Fuzzy *Confidence* (called *Min_FC*), the fuzzy rule *R* is considered acceptable. In this algorithm, a table structure, called FGTTFS, is implemented to create large fuzzy grids. This table includes the following substructures [111]:

- Fuzzy grids substructure (FG): each row represents a fuzzy grid, and each column represents a linguistic value.
- Transaction substructure (TT): each column represents a tuple t_p, while each element records the membership degree of t_p belongs to the corresponding fuzzy grid.
- Fuzzy support substructure (FS): stores the fuzzy support corresponding to the fuzzy grid.

Since linguistic variables are utilized in the FGBRMA, this algorithm is considered as an effective algorithm. This algorithm is well adapted for the thinking of human subjects and helps to increase the flexibility of systems. Some significant reasons why we utilize the FGBRMA algorithm instead of applying any modified version of the Apriori algorithm can be briefly found as follows.

1. This algorithm uses the linguistic hedge to change the meaning of the fuzzy terms. The meaning of the linguistic values of features can be changed by a linguistic hedge such as 'very' or 'more' or 'less'. It appears that these uses of the linguistic hedge will supply productively linguistic values, which will make the fuzzy association rules discovered from the database more flexible for the systems.

2. This algorithm defines the various number of linguistic values in each quantitative feature. The number of linguistic values defined in each quantitative feature need not be equal to *K*. In fact, decision-makers can determine possible linguistic values for one feature by using their preferences or domain knowledge.
3. This algorithm does not restrict the shapes of the membership functions defined in the features. Therefore, in this algorithm, we can straightforward refine the membership functions of linguistic values by using various machine learning techniques.

According to the significant reasons mentioned above, we can assert that based on the FGBRMA algorithm, it is feasible to develop an effective fuzzy associative classifier. On the other hand, database mining problems involving classification can be viewed within a common framework of rule discovery.

4.4.2 Fuzzy Clustering and Partition Generation

In this section, the fuzzy partition method for defining fuzzy membership functions is described. In fuzzy ARM, each feature is replaced by a various range of values using the fuzzy set theory. Generally, the membership functions (MFs) have a crucial impact on the association rules mining results, where various MFs extract different knowledge [112]. The fuzzy sets and their corresponding MFs are sometimes provided by an expert, in many evolutionary fuzzy association rules mining algorithms. However, this is not a suitable method to specify them by experts, because it takes much time. Furthermore, the sets may be changed in terms of the criteria that the user focuses on [113]. The foremost solution is to automatically determine a set of MFs to be applied for mining FARs. Based on the type of analysis carried out, both problems of mining FARs and MFs should be implemented simultaneously.

In this way, a rules mining algorithm is commonly divided into two stages [114]:

1. Determining MFs in which an EC algorithm is utilized to derive suitable MFs.
2. Mining FARs in which the best MFs derived by the first stage is applied to fuzzify the original transactions to fuzzy ones.

As mentioned above, the membership generation methods are generally established from expert knowledge or experimental approaches. Nevertheless, one way to characterize the membership functions of these linguistic values is based on the fuzzy c-means (*FCM*) clustering algorithm [110]. This algorithm is an unsupervised learning method to form data into clusters based on data similarity regardless of the target class information. In this research, we use the *FCM* clustering algorithm to determine MFs. In *FCM*, every data point belongs to every cluster to a certain degree μ in the range [0,1]. This algorithm tries to minimize the following objective function:

$$F_{obj} = \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2 \tag{4.1}$$

where μ_{ij} is the degree of membership of x_i in the cluster j, x_i is the *d*-dimensional measured data, c_j is the *d*-dimensional center of the cluster, and ||.|| is the norm that shows the similarity between any measured data and the center. The fuzziness parameter, *m*, is an arbitrary real number (m>1). The brief overview of this approach is as follows: Let *T* be a set of transactions, where t_1, t_2,…, and t_N are different crisp records, and $I = \{I_1, I_2, \ldots, I_n\}$ be the set of features. Also, let $FP = \{FP_1, FP_2, \ldots, FP_n\}$, where $FP_n = \{f1, f2, \ldots, fs\}$ be the set of fuzzy partitions of attribute I_n. In the proposed algorithm, the *FCM* clustering algorithm is employed as the ***FCM***(*I*) function. The pseudo-code of the ***FCM***(*I*) function is illustrated in Figure 4-12 [96]. In this research, the *FGBRMA* algorithm is employed as the ***FGBRMA*** (Dataset, *Min_FS*, *Min_FC*) function to extract fuzzy rules, and ***FCM***(*I*) function is called by the ***FGBRMA*** function for clustering. So, each feature is viewed as a linguistic variable, and the variables are divided into various linguistic terms.

The value of each feature is transformed into three linguistic terms (*Low*, *Medium*, and *High*), and triangular membership functions are used for each linguistic value to define the fuzzy membership function. In other words, each feature is divided into three sub-features with linguistic terms (i.e., the number of clusters is set to 3). In Figure 4-13, a typical feature I_n, is utilized to demonstrate the function of ***FCM*** and the creation of partitions. First, the values of

$t_p[I_n]$ (for each p; $1 \le p \le N$) are partitioned into three clusters using ***FCM***. Then for each generated cluster, the membership degrees of samples are exploited to fit a triangle. The fitted triangle represents the membership function corresponding to that cluster. As can be seen in Figure 4-13, the membership degrees of the samples to three clusters are plotted by diamond "♦", circle "•", and star "✳" symbols, respectively.

Get m (Fuzziness parameter)
    For each $I_n \in I$
        Get K (Number of Clusters)
        Until $\max\left\{\left|\mu_{ij}^{(K+1)}\right| - \mu_{ij}^{(K)}\right\} < \delta$
    For each $t_p \in T$ (p=1, 2, …, N)
        For each cluster j (j=1, 2, …, K)
            Calculate μ_{ij}

Figure 4-12 Pseudo-code of *FCM*(*I*) function.

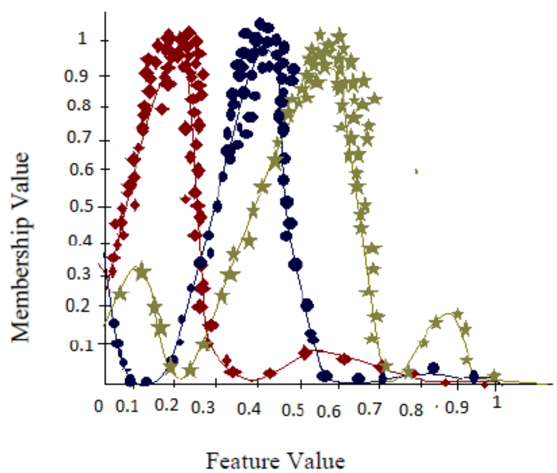


Figure 4-13 The fuzzy partitions of each feature in the example.

4.5 *Post-Processing of Pruning*

Computational complexity is a serious issue in association rules mining when data arrays contain a large number of rows and/or columns. The main reason for computational complexity is the inherent nature of a combinatorial explosion number of event associations. For real-world data, the search space is highly dependent upon the characteristics of the input data. Since the pattern associations are sparsely scattered in real data, the search space complexity increases rather fast in comparison with uniformly distributed in the hypothesis space.

Direct application of any rule mining algorithm to an image database would result in a large number of irrelevant associations. In general, association rules mining deal with a huge number of extracted rules causing slow training and extending the classification time. It could be an important problem in applications where fast responses are required. Therefore, this issue must be addressed. Furthermore, the redundant rules slow down the interactive process by the user, and it is easy to see that no matter how efficient the proposed framework is, the approach above will be slow for a large image dataset, and there will be a bottleneck for algorithm performance.

Association rules mining technique seeks interesting associations among features from massive high-dimensional categorical feature spaces. However, as the dimensionality gets higher, the data gets sparser which results in the discovery of a large number of association rules and makes it difficult to understand and to interpret.

Existing association rules mining algorithms rely only on frequency-based rule evaluation methods such as Support and Confidence, fail to provide sound statistical or computational measures for rule evaluation, and often suffer from low accuracy.

Some other techniques raise the argument that the pruning technique should be limited to only "negative" rules that lead to incorrect classification. These methods, called the database coverage approach, usually apply some parameters to restrict the rules. These parameters will be applied at the first stage of the algorithm and do not allow for rules generation.

Pruning non-essential rules without putting at risk the classification accuracy is important but very challenging. The pruning techniques must be used attentively since a wrong removal of rules may delete useful knowledge. Therefore, one solution is to apply a late database coverage approach, called lazy pruning, which occurs after the extraction of rules.

Recent studies indicate the lazy pruning techniques usually generate slightly higher predictive classifiers than those which apply the database coverage approach.

In this research, we have applied two pruning approaches based on "redundancy restriction" and "feature dimensionality reduction" to obtain a compact and light-weight associative classifier. Our pruning advantage over regular associative classifier methods is that it avoids the overfitting for the rules mining algorithm. In particular, *PrFA* is focused on producing individually significant rules and avoiding the exponential curse and bottleneck that would make the subsequent expert examination unappealing.

In this research, the main idea of pruning is to discover the best fuzzy rules that would make a considerable distinction between the classes. Our pruning techniques aim to discard irrelevant fuzzy rules to speed up the classification process. It is necessary to mention that pruning techniques must be used attentively since an extreme removal of rules may delete useful knowledge.

The pruning techniques that we apply in this research are the following:

- In the first type of pruning, we apply the chi-squared test to focus on the similar behaving features and testing strong dependence between them to reduce the amount of redundant and insignificant final rules.
- In the second type of pruning, we propose an engine based on a fuzzy associative feature selection method named Engine_*FAFS,* which is incorporated into the proposed framework to control the size of the feature subset of images.

At the end of the two steps of pruning, we obtain strong fuzzy rules, which can be applied for the classification task of unlabeled patterns. The set of fuzzy rules which survived after the two pruning steps will be applied as feature vectors to build an efficient and effective associative classifier. In the following, the details of all the mentioned pruned steps would be introduced.

4.5.1 *Chi-Square Test for Independence and Correlation*

The key stability of association rule mining is its integrity. This strength, however, comes with a major obstacle. The main problem emerging here is the huge number of final produced fuzzy rules. The experimental results show that most of these extracted rules are redundant or insignificant [12]. That is extraordinarily authentic for datasets whose image features are highly correlated.

To address this limitation, we need a correlation measure of the dependency among features. Based on the type of analysis carried out, an interesting idea for testing independence and correlation can be Chi-square test statistics. Therefore, a technique to overcome these problems by focusing on similar behaving features in this direction will be performed. The prominent difference in performance between this method compared to other algorithms is in the number of final produced fuzzy rules. This method decreases the number of final fuzzy rules and leading to a shorter running-time along with improved efficiency. In this way, we automatically execute mining procedures on the reduced dataset that produces a much smaller but richer set of fuzzy association rules, which has been advocated by experimental results.

In this research, the Chi-Squared test has been used to investigate the similar behaviour of features [115]. Chi-Squared test statistics (χ^2) is a widely used method for testing correlation and independence. The use of the Chi-Squared significance test for independence is more solidly founded in statistical theory. Besides, the Chi-Squared statistic simultaneously and uniformly considers all possible combinations of the presence and absence of the various features being examined as a group [116]. In this direction, our technique in detail will be performed by focusing on similar behaving features. The Chi-Squared test is based on the comparison of observed frequencies with the corresponding expected frequencies. It is applied to test the significance of the deviation from the expected values. The χ^2 value is determined as follows:

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} \qquad (4.2)$$

where *Obs* is the observed frequency, and *Exp* is the expected frequency. A χ^2 value of "0" implies the features are statistically independent. If it is higher than a certain threshold value, the independence assumption will be rejected. As mentioned earlier, three linguistic terms were determined for every feature and triangular membership functions were used for each linguistic value. So, contingency tables for features can be built. The rows and the column of these tables are the fuzzy partitions that belong to each of such features. The observed and the corresponding expected frequencies for all the tables will be established, and then the χ^2 value will be calculated based on the formula (6). Under the hypothesis of independence, the quantity χ^2 has a chi-squared distribution with degrees of freedom. This parameter refers to the number of values that are free to vary after restriction has been placed on the data. The degree of freedom is obtained as follows:

$$Degree\ of\ Freedo = (Row - 1) \times (Column - 1) \qquad (4.3)$$

If χ^2 is large, it indicates that the observed and expected numbers of observations differ greatly, and the hypothesis of independence should be rejected. The seven steps of the chi-squared test for correlation analysis are shown in Figure 4-14. Furthermore, the critical values of chi-squared distribution with K degrees of freedom are listed in Table 4-3.

1. State the hypotheses:
   H_0: There is no relationship between the two features (Independent)
   H_a: There is a relationship between the two features (Dependent)
2. Input the dataset with n fuzzy linguistic terms.
3. Construct contingency tables for each two fuzzy linguistic terms two by two.
4. Compute the χ^2 value.
5. Find the degrees of freedom.
6. Find the p-value (**Table 4-3**).
7. Once the p-value is known, compare it to α, the significance level:
   p-value$<\alpha \Rightarrow$ Reject H_0
   p-value$>\alpha \Rightarrow$ Fail to reject H_0

Figure 4-14 Seven steps of the Chi-Squared test for correlation analysis.

Table 4-3 The critical value of Chi-Squared distribution with K degrees of freedom.

	Probability of Exceeding the Critical Value				
K	**0.10**	**0.05**	**0.025**	**0.01**	**0.001**
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588

4.5.2 *Proposed Engine_FAFS (Fuzzy Associative Feature Selector)*

Feature selection is one of the most common methods used in machine learning and pattern recognition. Since this method aims to segregate the redundant and irrelevant features from the dataset, the dimension of the dataset will be decreased. Feature selection can be formulated as a multi-objective optimization problem, so a wide variety of optimization algorithms, such as evolutionary ones, can be used for this purpose [117]. Due to the photography technology advancement such that an image can be digitized at an extensive range of resolution levels, image data are typically unstructured and high dimensional. In this research, we applied the IBIS TRAX-3D scanning device, which enlarges the original image and presents delicate details into high-resolution views. This device provides a high resolution (depth map) in the order of 6 microns, lateral resolution in sub-micro, which is adequate to enable users to collect detailed data from tiny topographical peaks and valleys. To overcome the high dimensionality problem, data miners usually select only a minimal set of features that are significant for classifying the images. In recent years, different feature selection methods have been proposed as a pre-classification, which most of them return results in the form of a score for each feature. Although feature selection is a popular method to reduce the dimensions in data, it is still difficult for data miners to select features based on such a scoring scheme. However, various studies have presented different methods to solve the dimensionality problem by finding the best optimal set of features. The reduced set of features has been proven experimentally to enhance the performance of the learning process and also be able to build an accurate classification model. Generally, feature selection techniques are divided into three classes, including filter methods, wrapper methods, and embedded methods [118]. The main differences between the filter and wrapper methods for feature selection are that the filter methods [119] evaluate the relevance of features by their correlation with the dependent feature.

In contrast, wrapper methods [120] evaluate the usefulness of a subset of features by actually training a model on it. Embedded methods combine the qualities of filter and wrapper methods

[121]. These methods are implemented by algorithms that have their built-in feature selection methods.

Up to now, a few studies based on association rules mining have been proposed. Work-related to mining image content, we can mention the following. The authors in [122] exploited the association rule mining to calculate the weights to find the optimal features that are closely correlative with the class attribute. Although their method was interesting, it was quite complex and performance tests with cross-validation. In [123], a filter method for feature selection based on association rule mining has been proposed that determines a set of ARs that their consequence is the target class. Unfortunately, this method does not perform automatically and requires the user to select the features one by one based on the feature scores reported from the algorithm. The authors in [124], improved the algorithm by proposing a clustering technique to help users in finding appropriate groups of features. However, the clustering algorithm is still semi-automatic in the sense that users must determine the suitable number of feature clusters.

Based on the type of analysis carried out, although some studies based on association rules mining have been proposed, the potential of fuzzy association rules mining for feature selection methods is numerous and is still an open field of research.

In this subsection, we propose a method for feature selection by taking advantage of fuzzy association rules mining to improve the accuracy of image classification. The experimental results obtained corroborate the assumption that our proposed algorithm does not only can reduce the computational complexity of classification but also can effectively improve the precision of classification in image databases by dimensionality reduction via removing redundant features.

In this step, we propose an engine based on a fuzzy associative feature selection method named Engine_*FAFS,* which is incorporated into the proposed framework to control the size of the feature subset of images. Engine_*FAFS* discovers a minimal set of original features that preserves the ability of the images to improve the accuracy of classification.

Engine_FAFS is one of the most important components of the proposed framework. This engine aims to segregate the irrelevant and redundant features from significant ones. It identifies the relationships among features in ruleset R and then eliminates some unnecessary features. It finds the relationships among features in the ruleset and then eliminates some unnecessary features. Suppose rule r forms as $A \Rightarrow B$, where A is the antecedent, and B is the consequence. In this rule, itemset B depends on itemset A. Thus, all items in itemset B can be eliminated because they are redundant. The detail of the feature selection algorithm is presented as follows. The algorithm proposed here is a novel feature selection method based on image content, and it extends the technique that we proposed in [96].

- **Condition 1**. Given an itemset I and a transaction set T, where each transaction is a subset of I, a fuzzy association rule is said to be an "implication" of form $A \Rightarrow B$ denoting the presence of itemset A and B in some of the T transactions, assuming that A, $B \subset I$, $A \cap B = \phi$, and A, $B \neq \phi$. The returned rules have the format $A \Rightarrow B$. Suppose fuzzy rule "r" from the ruleset R forms as $A \Rightarrow B$, where A is the antecedent and B is the consequence.
- **Condition 2**. For fuzzy rule $A \Rightarrow B$, itemset B depends on itemset A. Thus, all items in itemset B should be eliminated because they are redundant.

The Engine_*FAFS* is started by calling the FGBRM(D, *Min_FS*, *Min_FC*) function. The flowchart of this function is depicted in Figure 4-15. This function takes the region dataset D[1], *Min_FS*, and *Min_FC* as the input arguments and returns the fuzzy association rule set. The fuzzy clustering method is needed to ensure the success of rule generation. Therefore, the FGBRM(D, *Min_FS*, *Min_FC*) starts the generation of partitions on features by calling the *FCM*(I) function (Figure 4-12). After calling this function, each feature is extended by its linguistic values. It is noted that by using the corresponding membership functions, defined with each linguistic value, the original region dataset D is changed into a fuzzy dataset D_f.

[1] The output of *Module* #1

With all the above conditions, we propose the following algorithm. Engine_*FAFS* utilizes *interesting*(r) Boolean function, for each fuzzy rule "r", to characterize whether the fuzzy rule r is interesting or not. The pseudo-code of the Engine_*FAFS* and *interesting*(r) function are shown in Figure 4-16. If interesting(r) function returns "TRUE," then the linguistic variables that only appear in the antecedent of the fuzzy rule "r" are extracted, and all of them are added to the F_o. Then, a simple deletion procedure is performed that cancels all the linguistic variables covered by the rule "r" from the ruleset R. At the final step, the feature set F_o will be the result of the feature selection process. In our proposed Engine_*FAFS* algorithm, δ is one of the most important parameters and the accuracy of the Engine_*FAFS* depends strongly on the precision of δ. Actually, this parameter controls the size of the feature subset, and it is clear that picking out the smaller values for δ causes the algorithm to extract the larger size of feature subsets. In this research, we apply the Particle Swarm Optimization (PSO) algorithm to characterize the optimum value of δ. The details about the application of the PSO algorithm are introduced in section 4-7.

4.6 *Associative Classifier*

Generally, classification is the task of learning a goal function, f, that maps each feature i to one of the predefined class labels j [125]. It is a form of pattern recognition, which is supervised learning, and its object can be text, image, audio, or video.

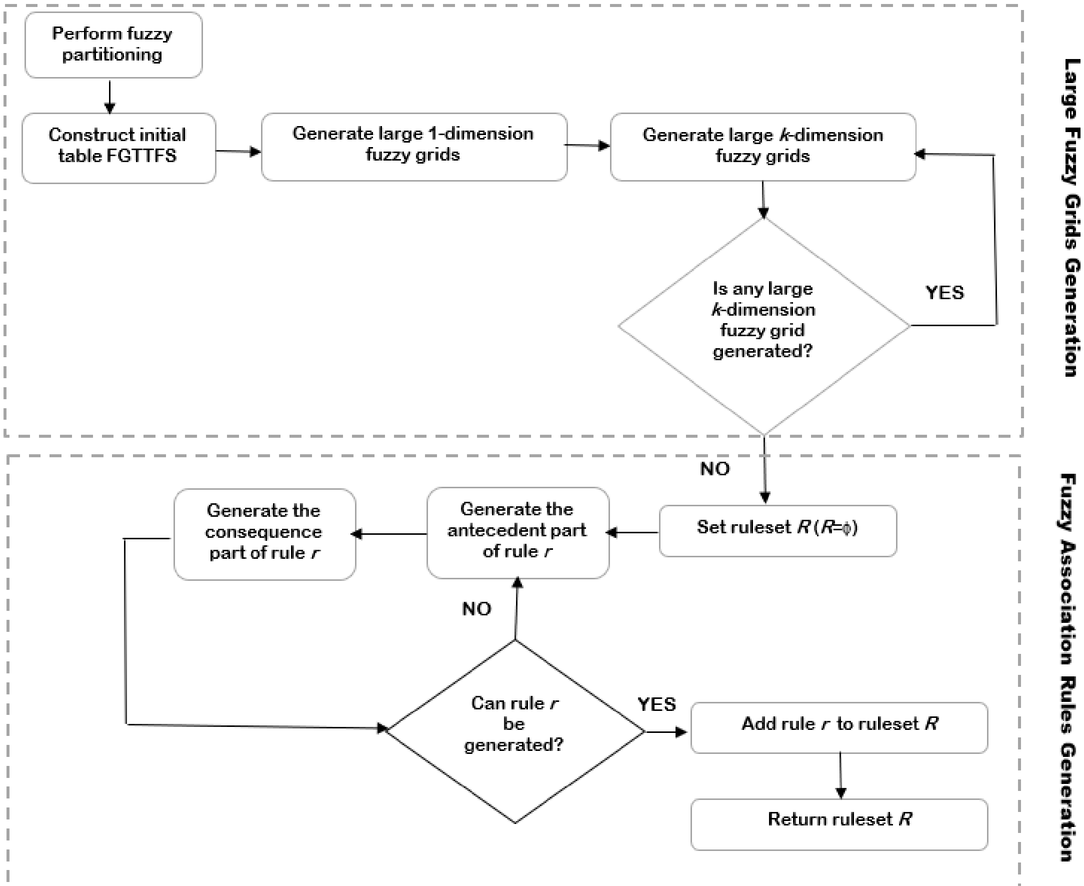


Figure 4-15 Flowchart of fuzzy grids–based rule mining function; *FGBRM(D, Min_FS, Min_FC).*

***"Engine_FAFS"* Algorithm**

Input 1. Dataset T of image transactions; ($I=\{i_1,\dots i_m\}$ be an itemset, each of which is an object in an image and T a fuzzy transaction set, in which each fuzzy transaction is a fuzzy subset of I)

Input 2. Size adjustment parameter (δ)

Output. Final feature subset (F_o)

```
F_o=ϕ
if (Ruleset R==ϕ) then break
else
    for each rule [r]∈R
        if interesting(r) then
            F_o= F_o+ (extract-linguistic-variable-in-antecedent [r]);
            Update R by deleting of all linguistic variables covered by the rule [r];
    end for;
end if;
Return F_o;
```

```
Boolean interesting (r)
Begin
if (Fuzzy Confidence [r]≥ δ)
    Return TRUE;
else
    Return FALSE;
end
```

Figure 4-16 The pseudo-code of the Engine_*FAFS* and *interesting* (r) function.

The classification is the process of recognizing a mapping function that allocates a new sample to its predicted class and demonstrates a relationship between input data and their class labels using a training data set. The process of classification is divided into two steps: the establishment of the classification model and the assessment of the resulting model. In the first step, a specific pattern that maps the training data to its class label will be detected. This pattern can be either a classification rule or mathematical formula. In the next step, this model is applied to predict the class of each testing data. The potency to predict the testing data is utilized as a benchmark to evaluate the precision of a model classification. Classification is aimed at the prediction of class labels, whereas association rule mining determines interesting relations between items in a transactional dataset [126]. One of the main advantages of an associative classifier is high-level interpretability while maintaining robust performance and managing missing observations [29].

The benefit of an associative classifier algorithm is that it can be represented in simple if-then rules, which makes it easy for the end-user to understand and interpret it.

Recently, the application of associative classifier has expanded to sentiment analysis in a plethora of decision-making systems. In recent years, a few studies have demonstrated that associative classifiers are superior to some traditional classification approaches, such as decision tree concerning prediction accuracy [86], [127]–[129]. Associative methods integrate association and classification functions in data mining. These methods include three phases:

- Rule Generation: It aims to extract all rules with high support, confidence, and length of the rule formed from the training data. The final rules set gets high priority to process than another rule.
- Building classifier: It aims to select the appropriate rules from the final rules set to construct a classifier.
- Predicting unknown class labels: It aims to assign the class labels of all cases in test data.

Associative classifiers incorporate classification and association rule mining to model construction, i.e. a classifier, in the task of classification. In general, ARM extracts an antecedent (or condition) and a consequent (or result) that have a conditional connection, such as "if a condition event takes place, a result event is likely." [29].

Associative classification is producing a rule-based classifier having extracted a set of association rules. In that manner, an association rule mining method firstly evokes a set of candidates for the final rule-based classifier. Then, these candidates are processed and combined to construct the final classifier. This step generally requires pruning techniques. After all, the classifier will be evaluated according to standard metrics on unseen test cases. Some considerable reasons that show the advantages of associative classifiers over other techniques can be briefly found as follows:

- They can exploit more correlations among the input features than traditional rule-based classifiers or decision trees [130]–[132].

- The expert can more easily interpret the results because the extracted rules illustrate a full interpretation of the prediction [133].

An associative classifier especially searches for a worthy set of rules that predicts the value of a target variable. Accordingly, contrary to ARM algorithms, it is conventional that all the rules have that variable in their consequent part. This aspect causes integrity in the associative classifier.

Although associative classification methods reported in the literature present several interesting aspects, they also suffer from some limitations. They do not deal with issues characteristic of the "Curse of Dimensionality". The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience. The expression was coined by Richard E. Bellman when considering problems in dynamic programming [134].

As mentioned earlier, association rules mining algorithms normally deal with a huge number of extracted rules causing slow training. Therefore, it is easy to see that no matter how efficient the proposed framework is, the approach above will be slow for a large image dataset, and there will be a bottleneck for algorithm performance. A pruned fuzzy associative classifier can be used to return, in addition to the result of the classification, the confidence in the result itself, to overcome the above deficiency.

Currently, the fuzzy sets theory has been successfully applied in various fields, especially in decision-making systems and classification problems. The advantage of the fuzzy set theory is dealing with uncertain and vague objects and is appropriate and enforceable because of no restrictions on the approximate description. In this section, we introduce a fuzzy classifier built on top of the pruned fuzzy rules. This part has the main contribution as follows:

a) Introducing a pruned fuzzy associative classifier for image classification
b) Compiling the precision of the proposed method and comparing its accuracy with other methods.

Our proposed associative classifier consists of two main steps:

1. Rule generation: The candidate rules that associate observations with class labels (class association rules; CARs) are extracted. It is a process of discovering frequent rule items (CARs) using the FGBRMA algorithm.
2. Rule Evaluation: *Support* and *Confidence* are used as measures to evaluate the significance of CARs in which redundant and less useful CARs are removed, involving rule ranking and pruning.

Candidate CARs are sorted according to particular criteria to make a preference that determines which rule will be included first in a classifier. Since class prediction for a new observation naturally follows by aggregating the results of several top-ranked CARs, the rule ranking process to measure the importance of CARs is necessary. It is noted to mentioned that, the accuracy of AC will progress if the evaluation of CARs improves, and effective CARs with high predictability will be selected. In this research, we have proposed an engine based on a fuzzy associative feature selection method named Engine_*FAFS*, which was incorporated into the proposed framework as a pruning method to control the size of the feature subset of images. Furthermore, the Chi-square test ($\chi 2$) approach to pruning rules has also been applied. Our framework has used this test to determine whether the antecedent and the consequent of a rule are correlated and whether the rule will be removed. If a positive correlation exists, the *PrFA* stores the rule in a classifier, otherwise it removes the rule from consideration.

4.6.1 *Pruned Fuzzy Associative Classifier (Classifier$_{PFA}$)*

The integration of supervised classification and association rules to build a classifier is not new. However, the main reason why we adopted the fuzzy association rules mining technique in developing our classifier is to perform a TOP-DOWN search to extract relationships among the features of coin images. The proposed classifier extracts semantic information embedded in image data. It identifies the images via their concept formed by objects with their relation.

Associative classifiers are especially appropriate to applications where the model may assist the domain experts in making decisions. The classification of coin images is a difficult and often computationally overwhelming task. Moreover, these images contain several correlated features, often referring to experts, which, when mined and exploited, can lead to superior classification.

In recent years, different methods based on the image for counterfeit coin detection have been proposed. Previous studies have developed heuristic/greedy search techniques to classify coin images. Apart from their usefulness, those methods focus basically on low-level features and neglect the conceptual association among the objects in the images.

From the different methods utilized for counterfeit coin detection, fuzzy association rules mining has been selected because it works remarkably well in discovering the implicit information from the images. Knowledge extracted from the image mining domain can be more descriptive compared to other areas.

Among the advantages of associative classifiers, we highlight two major ones:

1. Training sets with high dimensionality can be handled with ease and no assumptions are made on dependence or independence of features;

2. The classification model is a set of rules easily understandable by humans;

According to the underlying semantics, images can be classified in two ways: i) Classify by some main object, and ii) Classify by multiple objects with their relations.

Association rules mining for image classification is an effective way to deal with the second type of images.

For a large real-life image database, an image may contain several significant objects that form the concept of the image together. Furthermore, these two types of images may even exist simultaneously.

In our research for counterfeit coin detection, there are different types of features like stroke width, contour smoothness, lettering height, lettering width, relative angle, and relative distance. These features form the concept of the coin image together.

Hence, our method for classification focuses on the fuzzy association rules mining that aims at reducing the semantic gap between high-level human perception of images and low-level image feature representation. Nevertheless, none of the existing image classification methods can correctly classify both types of images at the same time although they can handle one certain type of image well individually.

From the different methods utilized for counterfeit coin detection, fuzzy association rules mining has been selected because it works remarkably well in discovering the implicit information from the images. knowledge extracted from the image mining domain can be more descriptive compared to other areas. In this section, we present a pruned based fuzzy associative classifier (classifier$_{PFA}$), which is built based on a training image dataset and used to associate class labels with previously unseen samples of images. In this approach, a set of fuzzy rules among the feature values and class labels are first extracted from the entire training dataset at once, and then the rules filtered after the pruning phase are integrated so that the collective fuzzy rule set has high predictive power to build the classifier$_{PFA}$. For the proposed classifier, we will apply the rigorous constraints so that only the fuzzy rules that can be used further for classification will be extracted.

An association rule used for classification is an "implication" of the form $A \Rightarrow C$, where itemset A is a non-empty subset of all possible items in the database and C is a class label. The pattern classification includes allocating a class C from a predefined set $C \in \{C_1, C_2, \ldots, C_M\}$ of classes to an unlabeled pattern. We define the fuzzy associative classification problem using a training dataset T with n distinct features $F_1, F_2, \ldots, F_n$, and $C \in \{C_1, C_2, \ldots, C_M\}$, a list of classes.

Given the transaction model described above, we are interested in a set of fuzzy association rules with n distinct features of the form $F_1 \wedge F_2 \wedge \ldots \wedge F_n \Rightarrow C_j$ where $F_i (1 \leq i \leq n)$ is the selected feature, and C_j is a category (G or F). In this way, the antecedent part of a rule is a subset of image features, and the consequent part is a class label. In other words, we constrain the fuzzy association rules such that the antecedent of the fuzzy rules is composed of the conjunction of features from the coin image while the consequent of the rule always has the category to which the coin image belongs.

An associative classifier is a particular case of association rule discovery in which only the class feature is considered in the rule's right-hand side (consequent); for example, in a rule such as $A \Rightarrow C$, C must be a class Feature. An associative classifier is different from association rule discovery. The most apparent variation between association rule mining and associative classifier is that the latter considers only the class label in the rule's consequent. While the former authorizes multiple feature values in the rule's consequent. Table 4-4 demonstrates the main important differences between association rules mining and associative classifier. Since an associative classifier requires a subset of the discovered rules to predict the classes of new data objects, overfitting can be a critical issue. Overfitting often arises when the extracted rules execute well on the training data set and badly on the test data set. It can be due to several reasons, such as a small amount of training data set or even noise.

Table 4-4 The main important differences between associative classifiers and association rules mining.

Associative Classifier (AC)	**Association Rules Mining (ARM)**
There is only one feature (class label) in the consequent part of a rule.	There could be more than one feature in the consequent part of a rule
The main objective is to construct a classifier that can predict the classes of test data.	The main objective is to discover associations among items in a transactional database
A class must be given (supervised learning)	No class attribute involved (unsupervised learning)
Overfitting is a serious issue	Overfitting is usually not an issue

Given an image to classify, the features of the test image are automatically derived, and a transaction is created by using the method discussed in the processing module (*Module*#1). Once the features of the test image have been discovered, they are taken and compared with the antecedent parts of all rules. The matching procedure for the test image would activate a list of appropriate rules that have the same features in their antecedent parts. From the activated rules, those that do not have a class in their consequent part are discarded, and the remaining rules are moved to a box which is called *unified-ruleset*. The rules that are moved to the unified-ruleset have high predictive power. In this step, all rules in the *unified-ruleset* are first grouped into *M*[1] boxes by the class that appeared in the consequent part, and then each box receives a vote based on a rule ranking procedure. The box which has a higher rank would indicate the most significant category that should be attached to the new image as a label. Figure 4-17 illustrates the pseudo-code of the classifier$_{PFA}$. To form an efficient associative classifier, we apply a procedure for voting consists of rule ranking in the limit given by the following metrics:

$\boldsymbol{R_G}$: A similarity metric to rules related to Genuine-box
$\boldsymbol{R_F}$: A similarity metric to rules related to Fake-box
$\boldsymbol{N_G}$: Number of rules in the Genuine-box
$\boldsymbol{N_F}$: Number of rules in the Fake-box
$\boldsymbol{Avg_G(Min\_FC)}$: The average of *Min_FC* for all rules in the Genuine-box
$\boldsymbol{Avg_F(Min\_FC)}$: The average of *Min_FC* for all rules in the Fake-box
$\boldsymbol{Label\ (Max(R_G, R_F))}$: $Returns$ "Genuine" $if\ Max(R_G, R_F) = R_G;\ Otherwise\ returns$ "$Fake$"

"Classifier $_{PFA}$" Algorithm

Input1. A new image to be classified; Pruned based fuzzy rules; *Min_FC*
Output. Class label attached to the new image

Method:

1. *"unified-ruleset"* = ∅
2. **EXTRACT** features of new image
3. **FOREACH** fuzzy rule **DO {**
4. **COMPARE** (features of new image) *AND* (antecedent part of the rule)
5. **If** [(features) *MATCHES* (antecedent part of the rule)] *AND* [(consequent part of the rule) *BELONGS TO* ((*"G"*) *OR* (*"F"*))] **THEN** *MOVE* rule to *"unified-ruleset"*
6. **}**
7. **DIVIDE** *"unified-ruleset"* in subsets by classes (*"G"*, *"F"*)
8. $R_G = N_G \times Avg_G(Min\_FC)$
9. $R_F = N_F \times Avg_F(Min\_FC)$
10. **LABEL** the test image by the class with $Label\ (Max(R_G, R_F))$

Figure 4-17 The pseudo-code of the classifier$_{PFA}$.

4.6.2 *Classifier$_{PFA}$ with Rejection Option*

This subsection represents a two-class (or binary) classification of elements X in $\mathfrak{R}^k$ that authorize a reject option. Based on the pair of features (X, Y) with $X \in \mathfrak{R}^k$ and $Y \in \{0, 1\}$, we consider the

[1] *M*: Number of classes

Classifier$_{PFA}$ that renders three possible outputs: 0, 1 and "*Reject*". "*Reject*" displays uncertainty and is to be applied for a few instances that are difficult to classify automatically.

The reject option presented by Chow [135] results in decisions not being taken for samples for which confidence is lowest, to decrease the likelihood of error. For pattern recognition applications that require high classification reliability, the reject option turns out to be very advantageous to secure against extreme misclassifications. In some application areas like fraud detection, healthcare, or biomedical data analysis, the certainty of the prediction is nearly as serious as the class label itself.

Generally, the primary intention of supervised pattern recognition is to classify the majority of future observations automatically. However, authorizing the reject option ("Unknown") besides taking a hard decision (0 or 1) is of appreciable emphasis in practice, for instance, in case of medical diagnoses or fraud detection. Nevertheless, this option is adequately ignored in the literature.

In this subsection, the reject option is presented by allowing the classifier$_{PFA}$ to withhold the decision of assigning a test image to any subset of categories (Genuine or Fake), for which the decision is considered not sufficiently reliable.

Given a test image i, the classifier$_{PFA}$ can automatically label i as belonging or not to any subset of the genuine or fake categories, while it rejects i from the remaining categories, i.e. no decision is taken about these latter categories. When the reject option is applied, it is straightforward to consider the costs of rejections in the determination of the expected risk. In this case, it turns out that minimizing the expected risk is equivalent to reconnaissance the best trade-off between the misclassification and rejection rates, depending on the corresponding costs [135], [136].

Given the above difficulties in defining the cost of rejections, we consider simply the rate of rejected decisions, i.e. the percentage of rejected category images over all the test image datasets, which will be denoted in the following as the "reject rate". Rejected patterns must then be categorized by a variant classifier, or by a human expert. It needs to discover a trade-off between the achievable reduction of the cost due to classification errors, and the cost of managing rejections. The problem of defining a reject option has been tackled only occasionally in literature.

Based on the type of analysis carried out, no work in the literature addressed the problem of defining a specific rejection technique for fuzzy associative classifiers.

Definition

The standard approach to rejection in pattern recognition is to approximate the class conditional probabilities and to reject the most unreliable objects, that is, objects that have the lowest class posterior probabilities. Consider our binary classification problem in which each test image belongs to one of two categories (Genuine or Fake). If the classifier_{PFA} is not sufficiently accurate for the task at hand, then a reject option approach can be utilized. The classifier_{PFA} can reject an instance if the prediction is not sufficiently reliable and falls into the rejection region. There is a relationship between error and rejection rate: according to [135], the error rate reduces monotonically with the growing rejection rate. Unfortunately, in real-world applications, this relationship is affected by considerable estimate error. In classifiers with a reject option, the key parameters are the thresholds defining the rejection regions. In this research, we do not deal with the problem of the optimal tradeoff between error and rejection. Our reject option deals with a single threshold and with binary classification only. It relies on the reject option with one global threshold.

As mentioned in earlier, once the features of the test image have been discovered, they are taken and compared with the antecedent parts of all rules. The matching procedure for the test image would activate a list of appropriate rules. These rules are then moved to a box (unified-ruleset). All rules in the unified-ruleset are first grouped into two categories (G or F) by the class appeared in the consequent part, and then each group receives a vote based on a rule ranking procedure. The group which has a higher rank would indicate the most significant category that should be attached to the new image as a label.

In this step, the reject option is applied based on the assessment of the reliability of the classification procedure utilizing a reliability evaluator. Once a reject threshold σ has been determined, an instance is rejected if the corresponding value is below σ. It is determined based on the output of the classifier; for a given sample, if the rank of the output is greater than a threshold σ, the classification will be accepted, otherwise, the sample will be rejected. In Figure 4-18, we modify the pseudo-code of the classifier$_{PFA}$ for the reject option which operates based on the reliability evaluation. The optimal value of the reject threshold (ρ) is determined through a training phase. Since the reject threshold demonstrates the minimum tolerable classification reliability level, when its value changes the reject option becomes more or less severe [137].

"*Classifier* $_{PFA}$" with Rejection option

Inputs. A new image to be classified; Pruned based fuzzy rules; *Min_FC; ρ*
Output. Class label attached to the new image

Method:

```
"unified-ruleset" = ∅
EXTRACT features of new image
FOREACH fuzzy rule DO {
   COMPARE (features of new image) AND (antecedent part of the rule)
     If [(features) MATCHES (antecedent part of the rule)] AND [(consequent part of the rule) BELONGS TO
     (("G") OR ("F"))] THEN MOVE rule to "unified-ruleset"
     }
IF "unified-ruleset" ≠ ∅
 {
    DIVIDE "unified-ruleset" in subsets by classes {"G","F"}
     R_G = N_G × Avg_G(Min_FC)
     R_F = N_F × Avg_F(Min_FC)
     IF (|R_G − R_F| ≥ ρ)
         {
           LABEL the test image by the class with Label (Max(R_G, R_F))
     ELSE
           LABEL the test image by "Rejected"
         }
ELSE
     LABEL the test image by "Rejected"
}
```

Figure 4-18 The pseudo-code of the classifier$_{PFA}$ with the rejection option.

4.7 *Optimization for the PrFA Framework*

Due to the essence of high-dimensional spaces, ARM is hard to solve. Hence, common techniques cannot provide sophisticated solutions, which have resulted in increased trendiness of non-exact ingenious optimization approaches and evolutionary computations (EC). Generally, there is a classification for EC algorithms: nature-inspired and non-nature-inspired algorithms [138]. Nature-inspired algorithms are categorized into different groups; one of the most important of them is Bio-inspired algorithms. Bio-inspired algorithms are influenced by biological science.

Swarm intelligence-based and evolution-based algorithms are two classes of bio-inspired algorithms; the origin of these approaches is the biological behaviour of natural objects. Swarm intelligence approaches simulate the collective behaviours of social swarms of birds or insects that live in a colony [1]. In the ARM field, many EC algorithms have been utilized to tackle the challenges of traditional ARM algorithms in terms of the number of rules extracted from large-scale datasets.

This section follows the Application of one of the Bio-inspired algorithms includes Particle Swarm Optimization (PSO) algorithm for the *PrFA* Framework.

Setting appropriate threshold parameters is a subtle task in our proposed framework. A high amount of thresholds terminates in a few rules while a low amount of them causes abundant (mostly redundant) rules. In addition to establishing minimum thresholds, the efficiency of the proposed framework depends precisely on dependent parameters. The *PrFA*'s parameters have a considerable influence on its performance, especially on the accuracy of classification.

PSO algorithm is mainly conducted by applying the key parameters that are significant to the convergence and efficiency. In this research, the algorithm proposed in [139] is modified to specify the optimum values for the *PrFA* framework automatically. In our proposed framework, four thresholds, namely, *Min_FS* (Minimum Fuzzy Support), *Min_FC* (Minimum Fuzzy Confidence), δ (to control the size of the features subset), and τ (to detect the blobs) should be optimized. Thus, this study presents a modified optimization method based on the algorithm proposed in [102] to specify the threshold values for the *PrFA* framework. Since the significant focus of this research is developing an image mining system on top of the fuzzy concept, we demonstrate the application of PSO for fuzzy association rules mining.

4.7.1 *The Application of PSO*

Kennedy and Eberhart first proposed the PSO algorithm in 1995 [140]. This algorithm is an evolutionary computational (EC) technique that simulates the nature of the particles in a swarm. The PSO algorithm prepares a population-based search procedure in which individuals, called particles, change their position (state) with time. In a PSO system, particles fly in a multidimensional search space. During the flight, each particle adjusts its position according to its own experience and neighbouring particle, making use of the best position encountered by itself and its neighbour. In this algorithm, each particle has a velocity and a position as follows [140]:

$$v_i(k+1) = v_i(k) + g_{1i}\big(P_i - x_i(k)\big) + g_{2i}\big(G - x_i(k)\big) \tag{4.4}$$

$$x_i(k+1) = x_i(k) + v_i(k+1) \tag{4.5}$$

where i is the particle index, k is the discrete-time index, v_i is the velocity of the ith particle, x_i is the position of the ith particle, P_i is the best position found by the ith particle (personal best), G is the best position found by a swarm (global best), and g_{1i} and g_{2i} are random numbers in the interval [0,1] applied to the ith particle. In our simulations, the following equation is used for velocity [141]:

$$v_i(k+1) = j(k)v_i(k) + \alpha_1\big[g_{1i}\big(P_i - x_i(k)\big)\big] + \alpha_2\big[g_{2i}\big(G - x_i(k)\big)\big] \tag{4.6}$$

in which $j(k)$ is the inertia function, and α_1 and α_2 are the acceleration constants. In this research, the linear decreasing strategy has been applied in which an initially large inertia weight (i.e., 0.9) is linearly decreased to a small value (i.e., 0.2) as follows:

$$j(k) = [j(0) - j(N_T)]\frac{(N_T - k)}{N_T} + j(N_T) \tag{4.7}$$

where N_T is the maximum number of time steps for which the algorithm is executed, $j(0)$ is the initial inertia weight, and $j(N_T)$ is the final inertia weight.

The PSO algorithm shares many similarities with EC techniques in general and GAs in particular. These techniques start with a group of the randomly generated population; all exploit a fitness value to measure the population. They all update the population and search for the optimum with random techniques. The principal discrepancy between the PSO approach compared with EC and GA techniques are that PSO does not have genetic operators such as crossover and mutation.

Particles update themselves with the internal velocity and also have a memory that is important to the algorithm. The information-sharing mechanism in the PSO algorithm compared with EC algorithms is meaningfully different. In EC approaches, chromosomes share information; thus, the whole population moves like one group towards an optimal area.

In the PSO algorithm, only the "best" particle gives out the information to others. Compared with ECs, all the particles tend to converge to the best solution quickly, even in the local version, in most cases. Compared to GAs, the advantages of PSO are that PSO is easy to implement, and there are few parameters to adjust [142].

There are two most momentous parts of this algorithm, including processing and mining. In the first part of the algorithm, the data are converted and stored in a binary format, and the search range of the particle swarm is set using the itemset range (*IR*) value calculated as follows [139]:

$$IR = \left[\log\big(mTransNum(m)\big) + \log\big(nTransNum(n)\big)\right]\frac{Trans(m,n)}{TotalTrans} \tag{4.8}$$

where $m \neq n$ and $m < n$. "m" represents the length of the itemset and $TransNum(m)$ means the number of transaction records containing m items. "n" is the length of the itemset, and $TransNum(n)$ means the number of transaction records containing n items. $Trans(m,n)$ means the number of transaction records mapping m to n items. *TotalTrans* represents the number of total transactions. By using of the *IR*, the itemset before the antecedent partition point is called "itemset *A*," whereas that between the antecedent partition and consequence partition points is called "itemset *B*."

In the second part of the algorithm, the PSO algorithm is utilized to mine the association rules. The steps of the PSO algorithm include encoding, fitness value calculation, population generation, best particle search, and terminate condition are explained below:

4.7.2 *Encoding*

This step is similar to the chromosome encoding of genetic algorithms. As mentioned in Chapter 3, Section 3.3.2, the intersection of the association rule of itemset A to itemset B ($A \Rightarrow B$) is empty.

So, items that exist in itemset A must not exist in itemset B, and vice versa. Therefore, both the antecedent and consequence partition of a fuzzy association rule are considered for the aim of chromosome encoding. In this research, each item is encoded into a string type chromosome by the corresponding order, which means that item 1 is encoded as "1" and item 2 is encoded as "2" and so on. An example of chromosome encoding has been shown in Figure 4-19. As can be seen in Figure 4-19, this chromosome has eight features, indicating that different fuzzy association rules can be generated. Thereupon, the *IR* value is applied to select the front and back part points of the chromosome. In this example, the *IR* value, $IR\{3 \rightarrow 7\}$, means that the minimal front partition point and the maximum back partition point are 3 and 7, respectively.

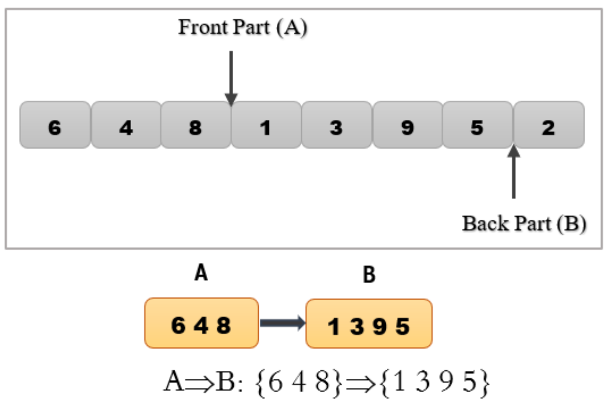


Figure 4-19 Chromosome encoding.

4.7.3 *Fitness Function*

The purpose is to extract the fuzzy association rules that both their fuzzy support and fuzzy confidence are larger than the other rules. So, the target function proposed in [143] is utilized to specify the fitness function as follows:

$$Fitness\ (k) = Confidence\ (k) \times log(Support(k) \times Length(k) + 1) \tag{4.9}$$

where $Fitness(k)$, $Length(k)$, $Support(k)$, and $Confidence(k)$ are the fitness value, length, fuzzy support, and fuzzy confidence of fuzzy association rule type k, respectively. The object is to maximize this fitness function. The larger the fuzzy particle support and fuzzy confidence, the greater the strength of association, meaning that it is an important fuzzy association rule.

4.7.4 *Population Generation*

The evolution commonly begins from a population of randomly generated individuals and happens in generations. This step is called initialization. In this research, particles that have larger fitness values are chosen as the initial particles.

4.7.5 *Search for the Best Particle*

Each particle adjusts its trajectory according to two factors, including "personal best" and "global best". The "personal best" is particle's initial position, which it has so far visited, and the "global best" is the best value obtained so far by any particle in the population. If the whole swarm is considered as a society, the "personal best" can be observed as resulting from the particle's memory of its past states, and the "global best" can be observed as resulting from the collective experience of all members of the society.

4.7.6 *Termination Condition*

Certainly, the consideration of a termination condition to complete particle evolution is vital. The termination condition occurs when the positions of all particles are fixed. In this research, the termination condition occurs after 150 iterations, and the evolution of the particle swarm is completed.

4.8 *Summary*

In this chapter, we have proposed a novel framework called *PrFA* for image-based counterfeit coin detection. Furthermore, a new algorithm named Engine_*FAFS* for feature selection and also an associative classifier named Classifier$_{PFA}$, have been proposed to create a robust counterfeit coin detector system. We reduced the amount of redundant and insignificant rules by focusing on pruning methods.

Chapter 5

Evaluation

This chapter reports on experiments, aiming to evaluate the performance of the proposed framework. In Sec. 5.1, we provide the experimental setup. In sec. 5.2, the augmentation methods utilized in this research are described. In sec. 5.3, the results of the optimization procedure are reported. In sec. 5.4, the analysis of the extracted fuzzy rules is presented. Sec. 5.5 describes the results obtained from the *PrFA* framework in comparison with some other recent methods in the field of counterfeit coin detection. In Sec. 5.6, the experiments carried out to test the classifier_{PFA} without and with a rejection option are presented. In pursuit of this evaluation, we also compare the results obtained by our proposed classifier with some other methods. Finally, this chapter is concluded in Sec.5.7 with a brief comparison of the *PrFA* framework with deep learning approaches.

5.1 *Experimental Setup*

The experiments of this research were conducted in the environment of Microsoft Windows 8.1-64 bit and hardware of the test environment consisted of an i7-4500U 4.2 GHz CPU (only one core was used), DDR3 6 GB RAM.

5.1.1 *Scanning Device*

In this research, we utilize a powerful scanning device – IBIS TRAX – in data sampling. As mentioned earlier, its patent belongs to Ultra Electronics Forensic Technology Ltd. Co in Montreal [19].IBIS TRAX-3D is a scanning device that can both capture 2D image and output 3D topography of exhibits. The scanning results are in a high resolution -- depth resolution in the order of 6 microns, lateral resolution in sub-micro, which is adequate to enable users to collect detailed data from tiny topographical peaks and valleys. Aside from that, there is a group of adjustable LEDs that are configured inside too. Control of LEDs provides us to have perspective images from different angles, thus allowing dynamic observation of the surface information. It supplies superior visualization via superposing the 2D texture image into the rendered 3D topography. Some of the main features of TRAX can be found below [19].

- 5-axis automated imaging. It supplies both rotational and lateral control and provides adequate motion over the exhibits. Therefore, it can digitize a precise and undistorted image.
- Wide field of view. It can decrease the distortion brought by stitching image patches.
- Orientation-independent annular lighting. It automatically removes operator variability and makes the device less prone to user error.
- Distortion-free orientation. It can supply a supreme view and excavate the fine details on the coin surface via a macroscopic level of ingredient both laterally and in-depth.

5.1.2 *Dataset Organization*

The dataset applied in this research consists of real data. The main reason why we study real coin images instead of any simulative data is to avoid insignificance, and we try to focus on the reliability of the discovered knowledge. It is important to note that having access to fake coins is a very difficult undertaking due to the legal issue, and the access to more fake coins is usually restricted. The Law Enforcement Office provided the coins used in this research.

Although research in coin authentication has been conducted for several decades, standardized datasets have not come into existence. The dataset that was usually employed in different literature consisted of images from various sources. However, the dataset utilized in this research includes real coin image data. Different human experts were also asked to identify coins instead of concentrating on a singular source to create a real dataset to ensure that the results are not biased due to a single source.

In this research, we created six types of coin image datasets using a powerful scanner. We scanned four different kinds of Danish coins: 20 Kroner 1990, 1991, 1996, and 2008 and two types of Chinese coins: half Yuan 1942 and one Yuan 1997. In these cases, counterfeit coins are very well-forged insofar, as detecting them is very difficult even to coin experts. The machine that was used for capturing the coins was a powerful scanning device named IBIS TRAX patented by Ultra Electronics Forensic Technology Ltd. Co in Montreal, Canada. IBIS TRAX-3D is a scanning device that can capture in 2D (with fast capturing) and 3D (with slow capturing). The machine has a built-in microscope, and a 5-group of adjustable LEDs are configured to view the object from different angles.

The precision of the scanner is about 6 microns resulting in a 3550×3550 image resolution with the gray level of 0 to 255 for JPG and 0 to 65535 for JPEG2000 images. In Figure 5-1, samples of genuine and fake coin images for the six different datasets resulted from scanning by IBIS TRAX are indicated.

Figure 5-1 Examples of genuine Danish and Chinese coin images: (a) Danish 1990, (b) Danish 1991, (c) Danish 1996, (d) Danish 2008, (e) Chinese 1942, (f) Chinese 1997; and (g) through (l) are the fake counterparts of the coins respectively.

5.1.3 *Memory Management*

As we mentioned earlier, a table structure, called FGTTFS, is implemented to extract fuzzy rules. Therefore, memory issues can arise during the construction of the FGTTFS table if the *Id* for each item consists of a very long string. A possible solution to this issue is to use a bitmap index for the *Id*s. When the large fuzzy grids are mined to generate rules, the relevant information is put in the rule table. Apart from this, the length of *Id* sets generated also needs adequate attention, and the sets may need splitting if they get too large to accommodate within limits imposed by database management systems on a single column size.

5.1.4 *Evaluation Metrics*

We will analyze the performance of the proposed framework according to standard evaluation metrics. Several experiments were conducted to evaluate the performance of the proposed framework as a tool for counterfeit coin detection. Before discussing the result of this study, it seems necessary to mention that although our proposed framework can easily classify multiple classes, the counterfeit coin detection is a binary problem, where only two class labels exist (Genuine and Fake). Therefore, the *Confusion Matrix* indicates the categories shown in Table 5-1, include the following.

- True-Negative (TN): The genuine test images are correctly classified.
- False-Negative (FN): The fake test images incorrectly classified, such as genuine test images.
- False-positive (FP): The genuine test images incorrectly classified, such as fake test images.
- True-Positive (TP): The fake test images correctly classified.

Some standard metrics are derived from the *confusion matrix* such as accuracy, recall, precision, the F-measure, and the AUC (Area Under the receiver operating characteristic Curve) measure that are the most common metrics to evaluate the performance of systems. These measures are calculated as follows:

Table 5-1 Confusion matrix.

Real Class	Predicted Class (Genuine)	Predicted Class (Fake)
Genuine	True-Negative (TN)	False-Positive (FP)
Fake	False-Negative (FN)	True-Positive (TP)

$$Precision = \frac{TP}{TP + FP} \tag{5.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5.3}$$

5.2 *Augmentation*

In this research, we exploit two different methods for data augmentation. We apply the first technique to generate genuine coins, and the second one to produce fake coins.

5.2.1 *Classic Augmentation of Genuine Coins*

A specific type of a genuine coin comes typically from a single source. Therefore, genuine coins are physically the same are different in quality. The images can also be different in the ways of capturing like rotation angle. Hence, augmenting train data for genuine class can be done by rotation and adding some noises. We add some Gaussian, and speckle noises to produce poor quality coins, and in some cases, we only rotate the images to produce good quality coins.

5.2.2 *Augmenting Fake Samples by Generative Adversarial Network (GAN)*

Unlike genuine coin images that follow the same patterns, fake ones are mostly from different sources. Therefore, classic augmentation may not be enough to generate fake samples. To address this issue to generate fake coins, we proposed a GAN based method. As shown in Figure 5-2, the general structure of the proposed GAN is rather standard with a minor change in the proposed GAN in [144]. In this research, the first block of the generator transforms the input into a 50×50

512-channel feature map where the resolution of the training images is selected, 200×200 in this research. The result of GAN for generating fake samples can also be enhanced. Figure 5-3 (a), (b), and (c) illustrate genuine coin images of 20 Kroner 2008, the generated image using the GAN, and the enhanced images of the generated images. Figures 5-3 (d) through (f) are also the same images for a Danish 1990 coin.

The number of augmented data for the training set is selected by applying a method suggested in reference [3]. Table 5-2 provides information about the datasets that we finally used for train, validation, and test processes. Here, the samples are randomly split into training (20%), validation (40%), and test (40%) sets before any augmentation. The validation set is used to avoid over-fitting and tuning the parameters of the proposed method. We keep most of the original samples for the test and validation processes and use augmentation for the training stage.

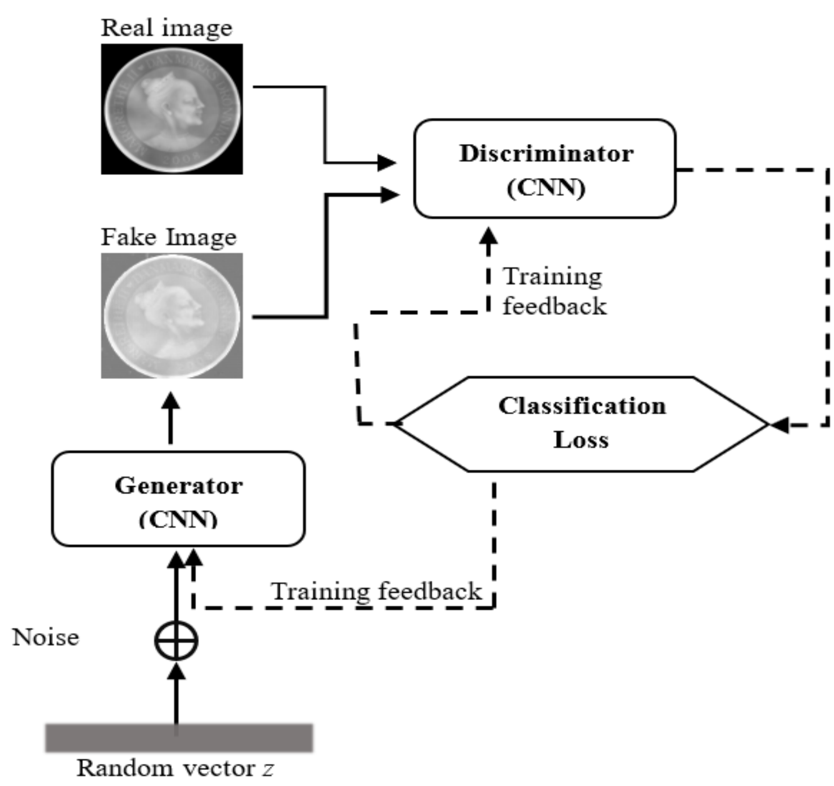


Figure 5-2 Structure of the GAN used for generating fake coins.

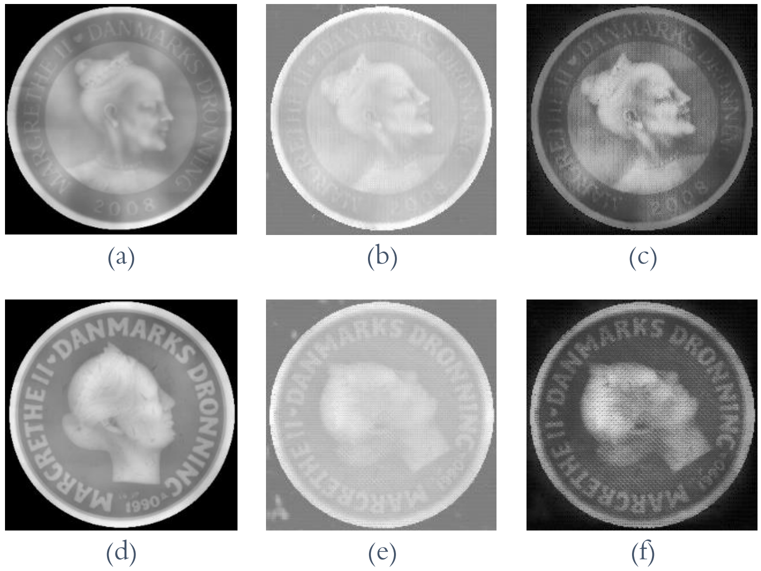


Figure 5-3 Generating fake samples by GAN. (a) a genuine 20 Kroner 2008-coin image, (b) the generated image by GAN, and (c) the enhanced image of the generated image and (d) through (f) are the same images for 20 Kroner 1990.

Table 5-2 Number of samples in training, validation, and test processes.

Datasets	Train before augmentation		Train after augmentation		Validation		Test	
	Genuine	Fake	Genuine	Fake	Genuine	Fake	Genuine	Fake
20 Kroner 1990	100	60	750	750	200	120	200	120
20 Kroner 1991	100	50	605	605	200	101	200	101
20 Kroner 1996	110	53	470	470	220	106	220	106
20 Kroner 2008	100	116	545	545	200	233	200	233
Half Yuan 1942	66	66	290	290	112	112	112	112
One Yuan 1997	50	51	250	250	100	102	100	102

5.3 *Optimization*

The *Support* threshold is the key to achievement in a fuzzy associative classifier. However, for specific application data, some rules with a large *Confidence* threshold are rejected since they do not have enough support. Classic algorithms use only a pre-determined *Support* threshold by the user to adjust the number of rules derived and may be unable to extract high *Confidence* rules that have low support. These algorithms typically tend to set a low support threshold, which may give rise to problems such as overfitting, generation of a large number of candidate rules with high CPU time, and storage requirements. To explore an ample search space and to discover as many high confidence rules as possible, we apply the PSO algorithm to characterize the optimum values of threshold parameters. As mentioned earlier, we have four parameters, namely, *Min_FS* (Minimum Fuzzy Support), *Min_FC* (Minimum Fuzzy Confidence), δ (to control the size of the feature subset), and τ (to detect the blobs) in which the values of them should be optimized. The optimized values of these parameters are specified automatically to satisfy the desired measures and to achieve the best results for the *PrFA* framework.

In this research, we capitalized on applying the PSO algorithm to determine the threshold parameters based on the best values automatically. The PSO algorithm is an evolutionary method that imitates the nature of the particles in a swarm. In pursuit of this optimization step, we compared the obtained results of PSO with the Genetic Algorithm (GA). In these experiments, both the PSO and GA were performed under the same conditions. In our simulations, different values are examined for GA parameters, such as crossover probability, mutation rate, and population size. Figures 5-4 and 5-5 clearly illustrate that the PSO algorithm outperforms the GA both in population size and the number of evolutions. It has been observed that the PSO algorithm arrives at its final parameter values in fewer generations, whereas the execution time increases with the number of population size and the number of evolutions for GA. Therefore, it can result from our simulations that the drawback of GA is its expensive computational cost.

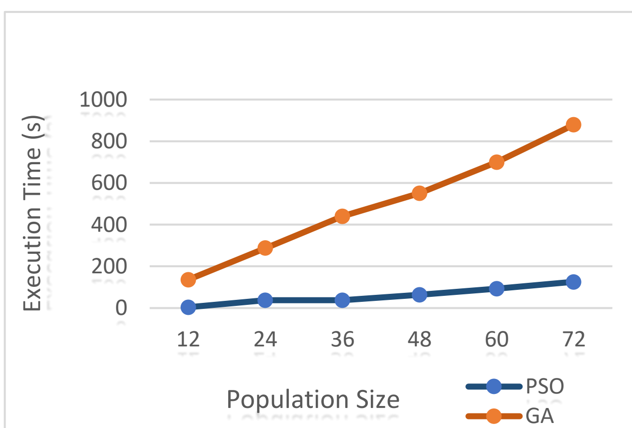


Figure 5-4 Population size and execution time for PSO and GA in the *PrFA* framework.

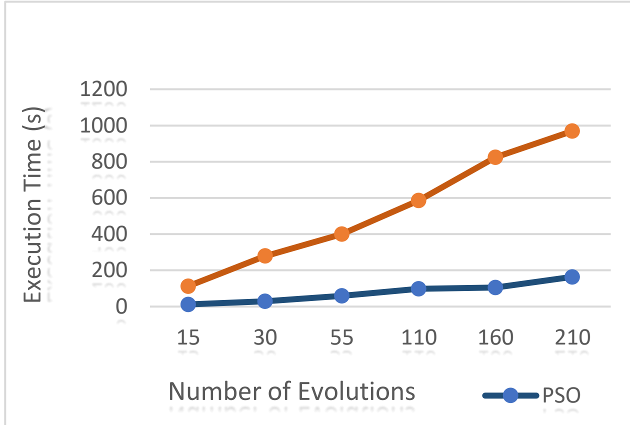


Figure 5-5 Number of evolutions and execution time for PSO and GA in the *PrFA* framework.

In these experiments, the PSO algorithm has the same effectiveness (finding the exact global optimal solution) as the GA but with significantly better computational efficiency (fewer function evaluations). The experimental results showed us PSO outperforms GA and can converge faster, which is very important, as the image dataset is usually large. Based on our experiments, it can be mentioned that the disadvantage of GA was its expensive computational cost, and the PSO algorithm had the same productiveness as the GA but with meaningfully higher computational efficiency. Therefore, we have conducted a series of experiments based on the PSO algorithm and have applied the results on the *PrFA* framework. Table 5-3 indicates the optimum values of the four parameters by using the PSO algorithm.

Table 5-3 Optimum values of threshold parameters specified by the PSO algorithm.

Parameter	**Value**
τ	0.5401
δ	0.4978
Min_FS	0.6311
Min_FC	0.8263

5.4 *Analysis of the Extracted Fuzzy Rules*

In this section, the fuzzy rule mining analysis will be discussed in detail. Sec. 5.6.1 compares the performance of the proposed classifier$_{PFA}$ by different blob detection algorithms to choose the more compatible ones with our framework. In pursuit of the analysis, we also provide the results obtained by the fuzzy association rules mining.

5.4.1 *Blob-Detection*

In the processing step, to extract features, we apply a blob detection method, which is incorporated into the proposed framework. Blob detection is one of the principal parts of the processing module of our research. Therefore, we accomplish an evaluation of three different techniques for finding blobs. We compare the performance of the proposed classifier$_{PFA}$ by three different state-of-the-art blob detection algorithms, including Blobword [72], MSER [145], and watershed [146]. This comparison is to find the best blob detection method, which is compatible with our framework. Figures 5-6, 5-7, and 5-8 show the size of the blob-set for each type of coin after removing the unwanted blobs using the Blobword method, MSER algorithm, and Watershed algorithm, respectively.

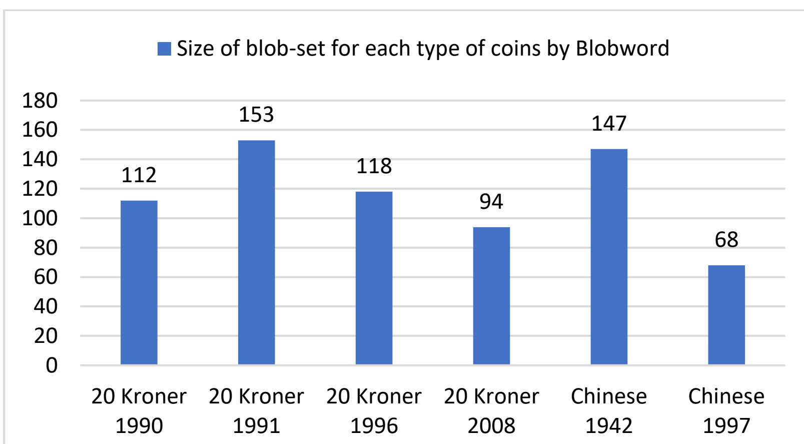


Figure 5-6 Size of blob-sets for all coin datasets after removing all unwanted blobs by the Blobworld method.

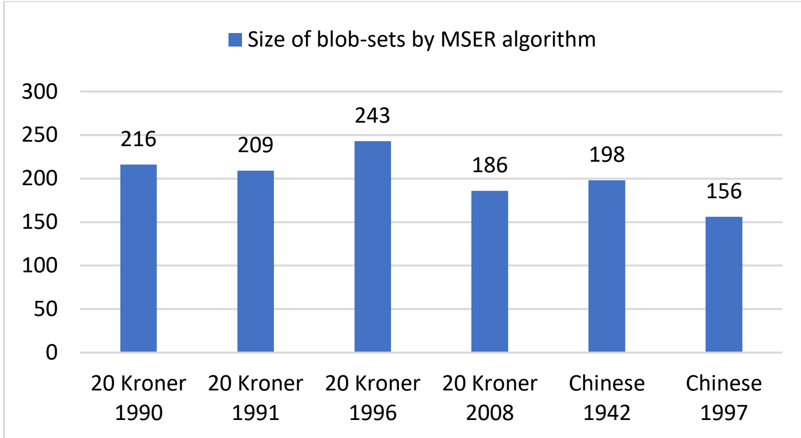


Figure 5-7 Size of blob-sets for all coin datasets after removing all unwanted blobs by the MSER algorithm.

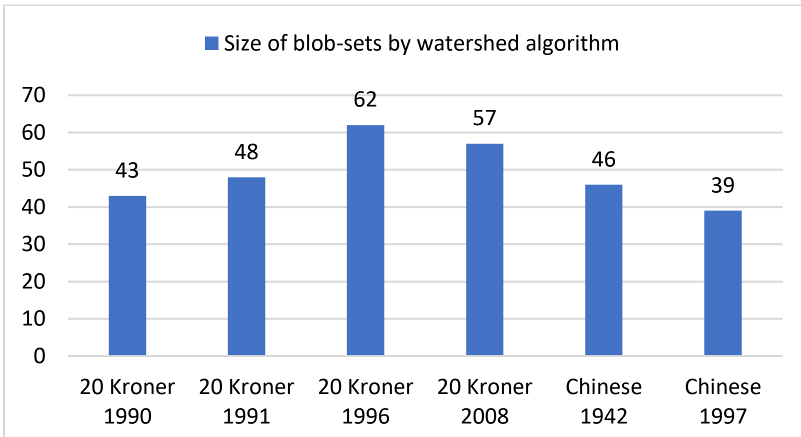


Figure 5-8 Size of blob-sets for all coin datasets after removing all unwanted blobs by the Watershed algorithm.

Table 5-4 indicates the effect of the blob detection method on the classification stage in terms of accuracy and AUC for three different mentioned method. As can be seen, the Blobword algorithm proposed in [52] outperformed other blob detection algorithms in terms of accuracy and AUC. According to results obtained from this comparison, we exploit the Blobworld method for the processing step.

5.4.2 *Fuzzy Rules Induction*

According to the obtained optimum values of the four parameters of the *PrFA* by using the PSO algorithm, the rules mining procedure will be automatically launched.

Table 5-4 Effect of blob detection on the classification stage in terms of accuracy and AUC.

Blob detection method	Metric	20 Kroner 1990	20 Kroner 1991	20 Kroner 1996	20 Kroner 2008	Half Yuan Chinese 1942	One Yuan Chinese 1997
Blobword [72]	Accuracy (%)	93.2	97.5	100	99.6	73.4	82.4
	AUC	0.955	0.991	1.000	0.999	0.954	0.903
MSER [145]	Accuracy (%)	88.0	87.1	97.7	97.0	71.6	79.8
	AUC	0.933	0.957	0.991	0.979	0.786	0.863
Watershed [146]	Accuracy (%)	68.0	66.8	73.4	86.3	69.7	62.6
	AUC	0.812	0.932	0.954	0.904	0.938	0.790

Starting with the *Min_FS* and *Min_FC,* the training dataset is mined, then a set of fuzzy association rules are found. The proposed framework extracts a set of fuzzy rules from the entire training dataset only once. These fuzzy rules are then arranged and applied as a classifier to test the validation dataset to avoid overfitting. The proposed classifier is a supervised learning method in which when the accuracy of the validation dataset is higher than a given accuracy threshold, the mining process is stopped. Otherwise, the amount of *Min_FS* and *Min_FC* will be changed, and the process is continued. The number of extracted rules is one of the most critical issues in the performance of the system. Neither its minimization nor its maximization makes the final set of fuzzy rules significant. That is since too small rule sets will frequently miss interesting relations in the data, and too large sets would need further efforts from the experts. In the ideal case, all the significant and relevant fuzzy rules are extracted and not any other, but the number is not known priorly.

On the other hand, the simple complexity is another essential issue in the rules mining. We consider the simple complexity as the number of conditions in the fuzzy rules.

In this research, a total of 4022 rules have been discovered from which 1022 rules are consisting of two elements, 936 rules with three elements, 1253 rules with four elements, and 811 rules with five elements. In this way, some of the mined rules as a sample are listed in Table 5-5. As seen in Table 5-5, $O_{i,j}$ is an object in the transaction dataset and integers show quantifying the occurrence of the object. We use G as a keyword for genuine images and F as a keyword for fake images.

Table 5-5 Some of the fuzzy association rules with Min_FS ≥ 0.6311 and Min_FC ≥ 0.8263.

Fuzzy Rules	*FS* Value	*FC* Value
$2O_{1,\ Low} \wedge 1O_{2,\ Medium} \Rightarrow 1O_{1,\ Medium} \wedge G$	0.7609	0.8902
$1O_{2,\ Medium} \wedge 1O_{3,\ Low} \Rightarrow 2O_{1,\ High} \wedge F$	0.9234	0.9433
$1O_{1,\ High} \wedge 2O_{2,\ Medium} \Rightarrow 2O_{2,\ Medium} \wedge G \wedge 1O_{3,\ Medium}$	0.7564	0.8345
$2O_{1,\ High} \wedge 2O_{2,\ Medium} \wedge 1O_{3,\ High} \Rightarrow 1O_{2,\ Medium} \wedge G \wedge 2O_{4,\ High}$	0.6734	0.9457
$1O_{2,\ Medium} \wedge 2O_{3,\ High} \wedge 2O_{4,\ Low} \Rightarrow 1O_{4,\ Medium} \wedge F \wedge 1O_{5,\ High} \wedge 2O_{6,\ Medium}$	0.7504	0.8654
$2O_{1,\ Low} \wedge 1O_{2,\ Low} \Rightarrow 1O_{1,\ High} \wedge F \wedge 2O_{4,\ High}$	0.8654	0.9901
$1O_{1,\ Medium} \wedge 3O_{2,\ Low} \Rightarrow 1O_{2,\ High} \wedge 2O_{4,\ Low} \wedge G \wedge 1O_{5,\ Medium}$	0.7547	0.8301
$2O_{2,\ High} \wedge 1O_{4,\ Medium} \wedge 1O_{6,\ High} \Rightarrow 2O_{3,\ Low} \wedge G \wedge 2O_{5,\ Medium} \wedge 1O_{6,\ Low}$	0.6401	0.8324
$1O_{1,\ Low} \wedge 2O_{3,\ High} \Rightarrow 2O_{2,\ Low} \wedge F$	0.9435	0.8652
$2O_{2,\ Medium} \wedge 1O_{3,\ Low} \wedge 2O_{4,\ Low} \Rightarrow 2O_{2,\ High} \wedge F \wedge 1O_{3,\ High} \wedge 1O_{5,\ Low}$	0.6723	0.9865
$1O_{3,\ Low} \wedge 1O_{4,\ High} \wedge 2O_{5,\ Medium} \Rightarrow 2O_{4,\ Medium} \wedge G \wedge 2O_{5,\ Medium} \wedge 1O_{6,\ High}$	0.7654	0.9603
$1O_{2,\ High} \wedge 2O_{4,\ Low} \wedge 1O_{5,\ Low} \Rightarrow 1O_{1,\ Medium} \wedge F \wedge 2O_{4,\ Low}$	0.6433	0.8650
$2O_{1,\ Medium} \wedge 1O_{2,\ Medium} \wedge 1O_{3,\ Low} \Rightarrow 2O_{3,\ Medium} \wedge F \wedge 2O_{4,\ Low} \wedge 1O_{6,\ Medium}$	0.7321	0.9753
$2O_{2,\ Low} \wedge 1O_{4,\ Medium} \Rightarrow 2O_{2,\ Low} \wedge F \wedge 1O_{4,\ Low}$	0.8609	0.9924
$1O_{2,\ High} \wedge 1O_{3,\ Low} \Rightarrow 2O_{1,\ High} \wedge 1O_{5,\ Low} \wedge G \wedge 2O_{6,\ Medium}$	0.6319	0.8740
$3O_{2,\ Medium} \wedge 1O_{3,\ High} \wedge 2O_{4,\ Low} \Rightarrow 1O_{1,\ Medium} \wedge F \wedge 1O_{3,\ Medium} \wedge 2O_{4,\ High}$	0.8427	0.9801
$2O_{2,\ High} \wedge 2O_{3,\ Low} \wedge 1O_{5,\ Low} \Rightarrow 1O_{2,\ High} \wedge G \wedge 1O_{5,\ Medium}$	0.7439	0.9351
$2O_{3,\ Low} \wedge 1O_{5,\ Medium} \Rightarrow 1O_{6,\ High} \wedge F$	0.6502	0.8617

5.5 *The Performance Comparison of the PrFA*

As the performance of the system is concentrated on counterfeit coin detection, in Table 5-6, we evaluate the proposed method by results in classifying the coins with the fake class. This table illustrates the performance of the proposed framework in terms of precision, recall, F-Measure, and Area Under the Curve (AUC). According to Table 5-6, the results to recognize the fake coins of dataset 1991, 1996, and 2008 are outstanding, while the results for other datasets are also desirable. It is interesting to note that, as expected, the capabilities of the proposed framework reveal the effectiveness of image mining techniques. Furthermore, the performance of *PrFA* when using the Engine_*FAFS* has been reported in this table. As can be seen, the obtained results have been compared with the "no feature selection" case. The obtained results prove that Engine_*FAFS* enhances the results. As is apparent in Table 5-6, the feature selection algorithm had adverse effects on the classifier when it was trained and tested by the Chinese Yuan 1997. The negative impact of the feature selection algorithm on this dataset could be due to a small number of the blob (See Figure 5-6) in the blob detection process.

Table 5-6 Performance comparison of the *PrFA* using Engine_FAFS in terms of precision, sensitivity, F-Measure, and AUC.

	Before feature selection				**After feature selection**			
Datasets	Precision	Recall	F-Measure	AUC	Precision	Recall	F-Measure	AUC
20 Kroner 1990	0.811	0.818	0.815	0.889	0.843	1.000	0.915	0.955
20 Kroner 1991	0.929	0.927	0.928	0.915	0.970	0.995	0.978	0.991
20 Kroner 1996	0.979	0.979	0.979	0.998	1.000	1.000	1.000	1.000
20 Kroner 2008	0.978	0.976	0.977	0.942	0.995	1.000	0.998	0.999
Half Yuan Chinese 1942	0.866	0.777	0.798	0.908	0.825	0.952	0.901	0.954
One Yuan Chinese 1997	0.807	0.651	0.691	0.783	0.565	0.520	0.542	0.710

In pursuit of this evaluation, we compare the results obtained by our proposed framework with four recent methods in the field of counterfeit coin detection. The same dataset and experimental conditions have been applied to all methods to make a fair comparison. Furthermore, the parameters have been selected according to the guidelines provided by the authors in the papers in which each method has been presented. The intention of this comparative analysis is purely to indicate the competitiveness of the proposed framework compared to other counterfeit coin detection methods referenced recently in the literature. However, the main goal is to prove a compressed system for image classification that can be considered as a knowledge attainment tool. Table 5-7 shows the accuracy of the proposed method in classifying both classes fake and genuine comparing with some other methods reported in [3], [15], [32], [40], [45]. Since the feature extraction algorithm in [15] is limited for the 20 Kroner 1991, 1996, and 2008 coins, it cannot participate in the comparison for the 20 Kroner 1990 and Chinese coins. As we can see clearly, our proposed method performs better than all the other recent methods. The results illustrate that the proposed method has reached a remarkable improvement in the classification of the coins, especially for the Chinese coins, Danish 1991, and Danish 1996 datasets. Regarding Table 5-7, albeit our method has classified Danish 1996 and 2008 coins more accurately comparing 1990 and 1991 coins, Table 5-6 demonstrates that this method performs superbly in recognizing the coins with the fake class in Danish 1990 and 1991 datasets.

Besides, we compare our proposed framework with the method presented in [3], in terms of the Receiver Operating Characteristic curve (ROC curve) for the fake coins. As the proposed method outperforms other methods for Chinese coin datasets dramatically, the ROC analysis will be applied to the Danish datasets. The ROC curves in Figure 5-9 (a) to Figure 5-9 (d) summarize the trade-off between the true positive rate (TPR) and false-positive rate (FPR) for the models using various probability thresholds. These figures illustrate that the proposed method has a remarkable superiority over the method presented in [3], especially for 20 Kroner 1990, and 1991. In Figure 5-9 (a), the ROC curve shows the proposed method is much more accurate than reference [3] in the trade-off between TPR and FPR in detecting counterfeit coins. The superiority of our proposed method is also repeated for other datasets with fewer differences (Figure 5-9 (b) to Figure 5-9 (d)).

Table 5-7 Comparison of the proposed framework with some other counterfeit coin detection systems in terms of accuracy in recognizing both genuine and fake classes.

Datasets	[3]	[15]	[40]	[45]	[32]	Proposed
20 Kroner 1990	92.9	NA	87.2	91.1	90.0	93.2
20 Kroner 1991	96.6	93.1	82.4	90.2	95.6	97.5
20 Kroner 1996	98.4	96.7	97.0	97.7	99.5	100
20 Kroner 2008	99.6	95.5	92.6	95.0	93.4	99.6
Half Yuan Chinese 1942	68.4	NA	70.5	66.3	65.7	73.4
One Yuan Chinese 1997	62.0	NA	64.2	65.5	68.2	82.4

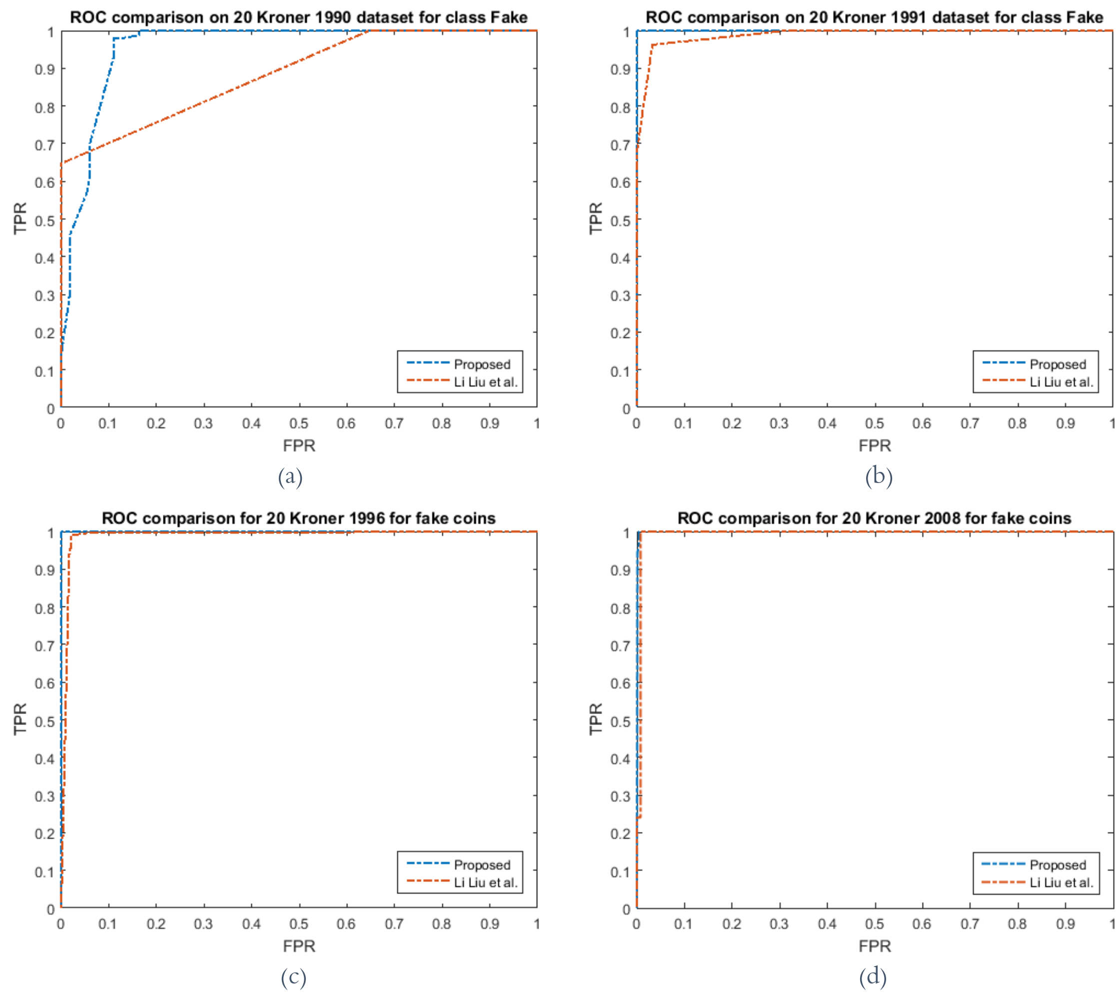


Figure 5-9 ROC comparison of four Danish coin datasets: (a) 20 Kroner 1990 dataset, (b) 20 Kroner 1991 dataset, (c) 20 Kroner 1996 dataset, and (d) 20 Kroner 2008 dataset.

5.6 *The Performance Comparison of the Classifier$_{PFA}$*

In this subsection, we compare the results obtained by classifier$_{PFA}$ with the ones achieved by four state-of-the-art rule-based classifiers [92], [147]–[149] feeding by our *final blob-set*. Besides, algorithms for class association rule mining in the literature belong to either the Apriori and complete methods family, or the evolutionary/swarm-based ones. Associative classification methods, either Apriori or evolutionary-based ones, are potential approaches for our domain, given that they internally apply class association rules. Thus, to evaluate the quality of the rules extracted by our approach, we compare them with those produced by Apriori and three representatives of associative classification. Furthermore, we investigate the problem of classification with the reject option, which is a technique applied to prosper classification reliability in pattern recognition systems.

In Table 5-8, we demonstrate the superiority of classifier $_{PFA}$ over four other state-of-the-art rule-based classifiers. As can be seen clearly in Table 5-8, the mentioned classifiers, especially C4.5 [147], have shown an acceptable result in terms of accuracy that proves the discriminating capability of our feature extraction method. However, the proposed classifier with the same input performed better than all the other classifiers.

Table 5-8 Comparison of the proposed classifier $_{PFA}$ with some other rule-based classifiers in terms of accuracy.

Method	*20 Kroner 1990*	*20 Kroner 1991*	*20 Kroner 1996*	*20 Kroner 2008*	*Half Yuan Chinese 1942*	*One Yuan Chinese 1997*
JRIP [92]	85.1	90.1	92.3	92.3	69.8	71.8
OneR [148]	84.8	89.9	92.4	91.9	70.1	71.3
PART [149]	89.2	86.4	93.0	94.6	66.5	69.4
C4.5 [147]	90.4	92.3	95.2	96.5	70.8	77.6
Proposed	93.2	97.5	100	99.6	73.4	82.4

Regarding the associative classification approaches, we also compare our proposed fuzzy associative classifier with those produced by Apriori and three representatives of associative classification:

- APRIORI [85]
- CBA (Classification Based Association) [150]
- CPAR (Classification based on Predictive Association Rules) [95]
- FARCHD (Fuzzy Association Rule-based Classification for High-Dimensional problems) [151]

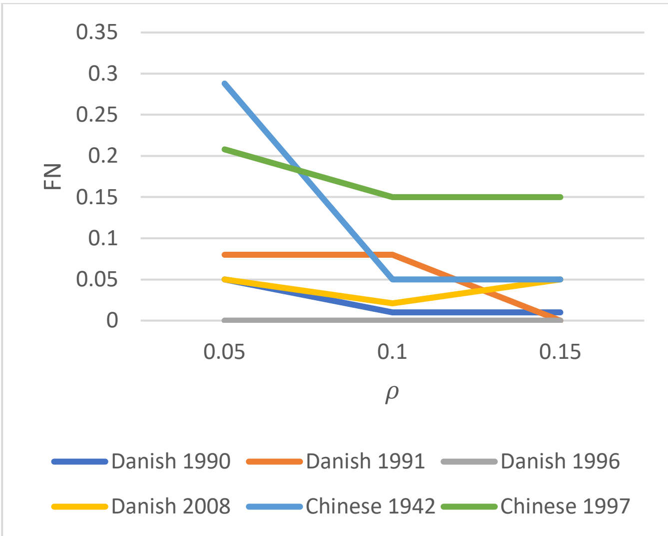
In Table 5-9, the number of final rules for different classifiers has been reported. According to Table 5-9, we observe that FARCHD and proposed classifier obtained results similar to each other, although some of the produced rules by FARCHD had confidence values less than 70%. During our experiments, we experienced that, although APRIORI achieved adequate results, it generated either invalid rules or too specific ones with very high confidence but minimal support. Invalid rules are usually due to the fact of having generated rules from a dataset with many missing values artificially imputed. In this case, the reason is partly due to the artificially imputed missing values, and partly due to the non-fuzzy interpretation we have made.

Nevertheless, the fuzzy interpretation did not serve for the APRIORI. As can be seen in Table 5-9, APRIORI has returned a few rules, in comparison with other methods. It means that APRIORI often either misses many possible interesting rules or excessively restricts the amount of the threshold parameters that should subsequently revise and interpret the rules. CBA returns only a few rules and CPAR, too many. Just FARCHD and our proposed framework have produced medium-sized sets. It is worthy of mention that, associative classifiers commonly seek a medium set of rules. Table 5-9 demonstrates that the *PrFA* generates an appropriate set of final fuzzy rules since it applies pruning techniques that guide the search and keep it focused. According to experimental results, we can assert that *PrFA* can extract the medium-sized of fuzzy rules, which is achieved with the power of fuzzy logic. Besides, it focuses on relevant features that avoid sporadic associations in the image data.

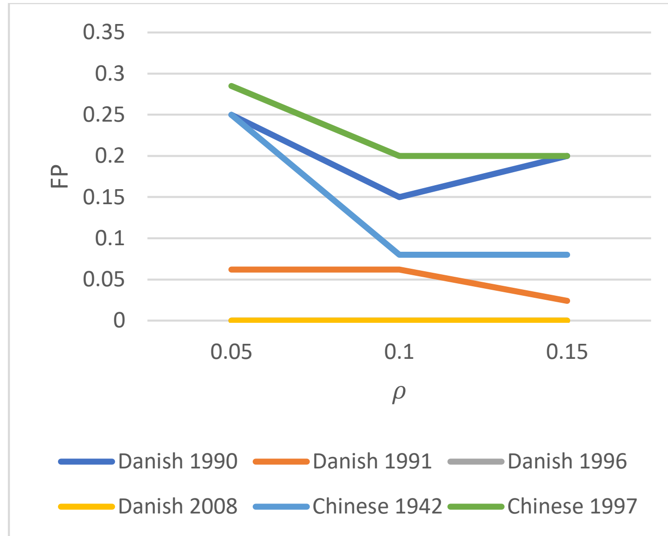
Table 5-9 Comparison of the *PrFA* with some other associative classifiers in terms of the number of final rule-set.

Method	Number of extracted rules
CBA [150]	3011
APRIORI [85]	2421
CPAR [95]	7387
FARCHD [151]	4739
Proposed	4022

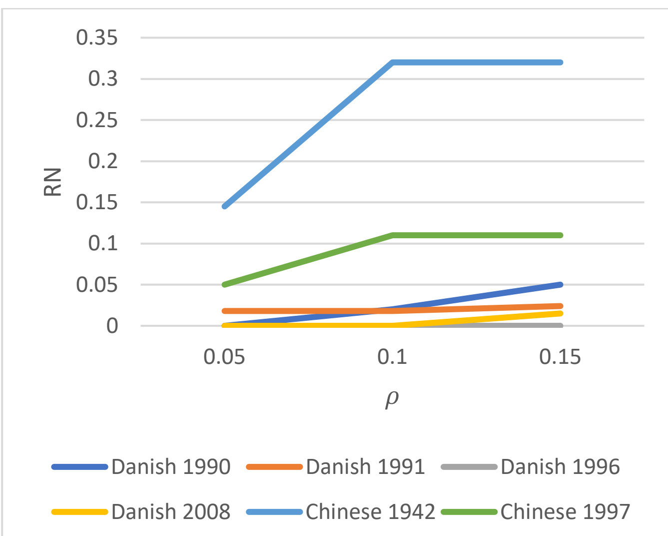
As mentioned earlier, to make our counterfeit coin detection system more reliable, we decided to supply the classifier$_{PFA}$ with a rejection option. It is necessary to determine the optimal value of the reject threshold (ρ). In this research, the optimal value of the reject threshold was specified through a training phase and the best value for ρ was selected 0.1. Figure 5-10 (a) and Figure 5-10 (b) illustrate that by increasing ρ, FP and FN errors are surpassed while the uncertain coins are rejected more shown in Figure 5-10 (c) and Figure 5-10 (d).



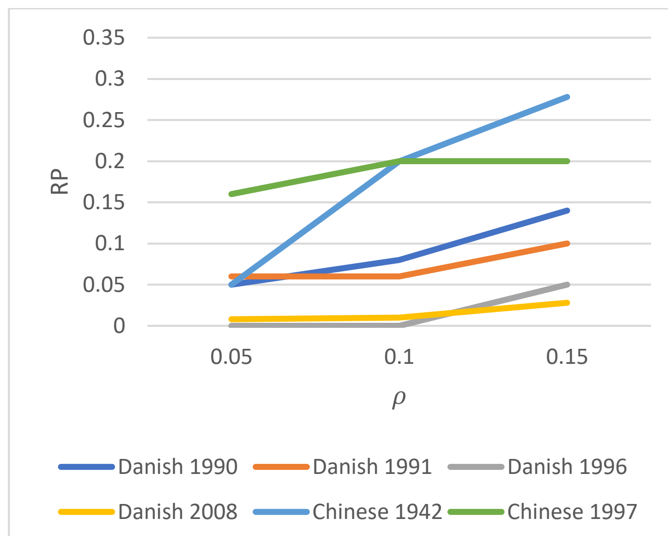
(a)



(b)



(c)



(d)

Figure 5-10 Selecting different ρ and its impact on: (a) False Negative, (b) False Positive, (c) Rejection Negative, and (d) Rejection Positive for all datasets.

To see the impact of the rejection option on reducing the errors and make the system more reliable, we provide TP, FN, RP, FP, TN, RN as well as Accuracy. Rejected Positive (RP), and Rejected Negative (RN) are defined by the percentage of counterfeit and genuine coins that are rejected, respectively. Table 5-10 illustrates the results for the metrics for all types of datasets. The accuracy in the classifier with rejection option is calculated by:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN + RP + RN)} \quad (5.4)$$

Table 5-10 Performance of the Classifier$_{PFA}$ with rejection option in terms of TP, FN, RP, FP, TN, RN and Accuracy.

Datasets	**Classifier$_{PFA}$ with rejection option**						
	TP	FN	RP	FP	TN	RN	Accuracy
20 Kroner 1990	0.91	0.01	0.08	0.15	0.83	0.02	0.87
20 Kroner 1991	0.90	0	0.10	0.02	0.95	0.02	0.92
20 Kroner 1996	1	0	0	0	1	0	1
20 Kroner 2008	0.98	0	0.01	0	1	0	0.99
Half Yuan 1942	0.75	0.02	0.23	0	0.71	0.29	0.73
One Yuan 1997	0.65	0.15	0.20	0.20	0.69	0.11	0.67

In order to compare the classifier with and without rejection option, we report the results of precision, recall, and f-measure in Table 5-11.

For all datasets in this experiment, the classifier with reject option has improved in terms of precision. Regarding the FP errors have dramatically been suppressed. However, the system with the rejection option did not perform better than classifier$_{PFA}$ without rejection. Although the FN errors have been reduced a bit, TPs have also been reduced with a larger ratio.

In conclusion, the system with the rejection option is more reliable when we are focusing on the detection of counterfeit coins. With this approach, we may have a large number of genuine coins rejected.

Table 5-11 Performance comparison of the proposed classifier$_{PFA}$ without and with a rejection option in terms of precision, recall, and F-Measure.

	Classifier$_{PFA}$ without rejection			**Classifier$_{PFA}$ with rejection**		
Datasets	Precision	Recall	F-Measure	Precision	Recall	F-Measure
20 Kroner 1990	0.843	1.000	0.915	0.858	0.989	0.919
20 Kroner 1991	0.970	0.995	0.978	0.974	1.000	0.986
20 Kroner 1996	1.000	1.000	1.000	1.000	1.000	1.000
20 Kroner 2008	0.995	1.000	0.998	1.000	0.978	0.989
Half Yuan Chinese 1942	0.825	0.952	0.901	0.903	0.937	0.920
One Yuan Chinese 1997	0.565	0.520	0.542	0.765	0.812	0.788

5.7 *Comparison of the PrFA with Deep Learning Approaches*

For Convolutional Neural Networks, we need thousands of images for training, validation, and test. The number of fake coins is strictly limited. Apart from this limitation, the similarity between fake and genuine coins in most of the types is extremely high where the automatic feature extraction may not be applicable. However, CNNs are very useful for coin recognition which is much easier than counterfeit coin detection.

To compare our proposed method with deep learning approaches, we applied three pre-trained CNNs VGG16, VGG19, and ResNet50 that have already been trained by the ImageNet dataset [152]. We unfreeze a few frozen layers at the top of the pre-trained networks to make fine-tuning possible. We also add a custom classifier on top of the pre-trained networks. We considered a flatten, a densely connected layer with 256 neurons and activation *relu*, and a dense output node with one neuron and activation sigmoid for the custom classifier. Table 5-12 illustrates that a fine-tuned VGG16 outperforms the other fine-tuned networks in terms of accuracy. However, the proposed method performs much better than the fine-tuned VGG16 tested by most of the datasets.

Table 5-12. Comparison of the proposed method with three state-of-the-art pre-trained deep learning CNNs in terms of accuracy.

Method	20 Kroner 1990	20 Kroner 1991	20 Kroner 1996	20 Kroner 2008	Half Yuan Chinese 1942	One Yuan Chinese 1997
VGG16	91.6	90.8	98.5	99.6	70.9	74.6
VGG19	86.7	92.4	96.3	94.5	72.5	69.4
ResNet50	90.5	90.0	93.5	94.8	68.5	71.2
Proposed	93.2	97.5	100	99.6	73.4	82.4

Chapter 6

Conclusions and Future Works

This chapter presents the main conclusions of the thesis and highlights future research work in related areas.

6.1 *Conclusions*

In this research, we introduced a novel framework called *PrFA* for image classification using a pruned fuzzy associative classifier. The problem domain for which we described a solution was the image-based counterfeit coin detection. In the proposed method, fuzzy association rules have been extracted to obtain frequently occurring local patterns in images. In this research, an image mining method using a blob detection technique has been introduced to find important information from the coin images. Besides, by using the proposed framework to a real coin image dataset, we indicated that it is feasible and has the potential to reveal significant patterns in the dataset. Dealing with quantitative values of image data has been one of the principal focuses of our method, in which the *PrFA* attempted to find the best membership functions derived by the Fuzzy C-means algorithm to fuzzify the quantitative transactions.

The defined strategies have been applied to establish the proposed scheme: a) detecting blobs to find objects in images automatically; b) creating the fuzzy transaction database by using the fuzzy partition method; c) applying pruning techniques based on redundancy restriction and feature dimensionality reduction; and d) applying a fuzzy associative classification method to classify the unlabeled samples. More in detail, we proposed a novel algorithm for feature selection in order to reduce the feature dimensionality. Our proposed Engine_*FAFS* algorithm can adequately utilize a large amount of data and extract important relationships among items and thus create a robust counterfeit coin detector system. Further, unlike previous methods proposed in the literature, which reduce the complexity by relying only on *Support* and *Confidence* parameters, we preserved the full power of fuzzy association rule mining by focusing on the similar behaving features and testing strong dependence between them.

Finally, we would like to remark that our framework has pointed out a research line not specially covered in the literature, the use of fuzzy set concepts for image-based counterfeit coin detection. Its advantage over standard associative classifier methods is that it focuses on the pruning techniques, so it avoids the bottleneck for algorithm performance. In particular, *PrFA* is focused on producing individually significant rules and avoiding the exponential curse that would make the subsequent expert examination unappealing. Furthermore, *PrFA* has allowed us to discover interesting rules among the massive amount of trivial and sporadic associations and has exposed the possibility of further work in this research line.

We believe that future advances in computer vision and image mining domains would benefit from adopting some of the suggestions provided to solve problems such as query expansion for image data, and auto-categorization of images that are perched at their intersection. We hope that our broad emphasis on pruning techniques and evaluation metrics to extract effective fuzzy rules will not only be helpful to the image mining community but will also stimulate the development of new solutions in this area.

6.2 *Future Works*

Directions for future work include the following. To the best of our knowledge, the framework proposed in this research demonstrates the first attempt of using fuzzy association rules mining for counterfeit coin detection. It is noted to mention that the accuracy of the proposed framework depends strongly on the exactness of the parameters of the Engine_*FAFS* algorithm and the parameters of the image miner module. This dependency causes an increase in search space. The results achieved by our framework, although already comparable to some other methods proposed recently in the literature, can be improved. As the proposed classifier extracts a set of fuzzy rules from the entire training dataset only once, one noticeable advantage of it is the time required for training (189.431 seconds), which is very low compared to other methods such as neural networks. Although the accuracy of the proposed method was the main concern of this research, it opens a promising research direction in time measurement for the future. Furthermore, applying the proposed framework for problems appearing in other domains, especially cooperation with medical stuff, would be outstanding.

Despite the advantages mentioned above, the proposed framework has some limitations. Although the proposed framework boosts the effectiveness of the discovered rules employed as discriminatory features and guarantees a higher level of efficiency in terms of accuracy, this advantage might not be enough to handle big data. Indeed, the process of the high dimensional data in the main memory is more difficult, and it would be precluded from storing all amounts of data in the main memory. This obstacle needs repetitious swapping operations, which can dramatically affect the execution speed of our proposed framework and also the computational time deteriorates considerably. Although this issue is universal to algorithms in data mining for big data, we aim to approach this problem for future work.

Despite the widespread achievement of ARM, there are many promising directions for additional research in the future:

a) Hybrid metaheuristic algorithms of single solution-based and population-based algorithms can be further considered for ARM.

b) In recent years, no progress has been obtained for ARM methods of high velocity of data, although it has acquired considerable improvement for large volumes of data.

c) There is a shortage of studies measuring the usefulness of evolutionary ARM algorithms under different aspects, which indicates an excellent opportunity for researchers.

d) Scalability is one of the most critical issues owing to an increase in the number of transactions in many real-world tasks. Recent improvements in ARM for large-scale global optimization represents that further studies on this area are needed.

Finally, we strongly believe that one of the most outstanding areas of research is to extract semantic information embedded in complex data. The field of mining complex data has a spacious variety of topics still not investigated, especially when coupled with image analysis. The prominent approaches to reduce the semantic gap between the high-level human interpretation and the low-level feature representation of images should be further constructed

References

[1] A. Telikani, A. H. Gandomi, and A. Shahbahrami, “A survey of evolutionary computation for association rule mining,” *Inf. Sci. (Ny).*, vol. 524, pp. 318–352, 2020, doi: https://doi.org/10.1016/j.ins.2020.02.073.

[2] H. Pan, J. Li, and Z. Wei, “Mining Interesting Association Rules in Medical Images,” in *Advanced Data Mining and Applications*, 2005, pp. 598–609.

[3] L. Liu, Y. Lu, and C. Y. Suen, “An Image-Based Approach to Detection of Fake Coins,” *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 5, pp. 1227–1239, May 2017, doi: 10.1109/TIFS.2017.2656478.

[4] N. T. Elkins, “A survey of the material and intellectual consequences of trading in undocumented ancient coins: a case study on the North American trade,” *Frankfurter Elektron. Rundschau zur Altertumskd.*, no. 7, pp. 1–13, 2008.

[5] K. Conlan, I. Baggili, and F. Breitinger, “Anti-forensics: Furthering digital forensic science through a new extended, granular taxonomy,” *Digit. Investig.*, vol. 18, pp. S66–S75, 2016, doi: https://doi.org/10.1016/j.diin.2016.04.006.

[6] G. Singh and K. Singh, “Digital image forensic approach based on the second-order statistical analysis of CFA artifacts,” *Forensic Sci. Int. Digit. Investig.*, vol. 32, p. 200899, 2020, doi: https://doi.org/10.1016/j.fsidi.2019.200899.

[7] M. X. Ribeiro, A. J. M. Traina, A. G. R. Balan, C. Traina Jr., and P. M. A. Marques, “SuGAR: A Framework to Support Mammogram Diagnosis,” in *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS’07)*, Jun. 2007, pp. 47–52, doi: 10.1109/CBMS.2007.101.

[8] O. R. Za\iane, M.-L. Antonie, and A. Coman, “Mammography classification by an association rulebased classifier,” *MDM/KDD*, pp. 62–69, 2002.

[9] A. Appice, M. Berardi, M. Ceci, and D. Malerba, “Mining and filtering multi-level spatial association rules with ARES,” in *International Symposium on Methodologies for Intelligent Systems*, 2005, pp. 342–353.

[10] M. X. Ribeiro, A. G. R. Balan, J. C. Felipe, A. J. M. Traina, and C. Traina, “Mining statistical association rules to select the most relevant medical image features,” in *Mining complex data*, Springer, 2009, pp. 113–131.

[11] S. Dua, H. Singh, and H. W. Thompson, “Associative classification of mammograms using weighted rules,” *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9250–9259, 2009, doi: https://doi.org/10.1016/j.eswa.2008.12.050.

[12] Z. Farzanyar, M. Kangavari, and S. Hashemi, "Effect of similar behaving attributes in mining of fuzzy association rules in the large databases," in *International Conference on Computational Science and Its Applications*, 2006, pp. 1100–1109.

[13] L. Liu, Y. Lu, and C. Y. Suen, "An Image-Based Approach to Detection of Fake Coins," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 5, pp. 1227–1239, 2017.

[14] M. Tresanchez, T. Pallejà, M. Teixidó, and J. Palacin, "Using the optical mouse sensor as a two-euro counterfeit coin detector," *Sensors*, vol. 9, no. 9, pp. 7083–7096, 2009.

[15] K. Sun, B. Y. Feng, P. Atighechian, S. Levesque, B. Sinnott, and C. Y. Suen, "Detection of counterfeit coins based on shape and letterings features," in *Proceedings of 28th ISCA International Conference on Computer Applications in Industry and Engineering, San Diego, USA*, 2015, pp. 165–170.

[16] S. Khazaee, M. S. Rad, and C. Y. Suen, "Detection of counterfeit coins based on modeling and restoration of 3D images," in *International Symposium Computational Modeling of Objects Represented in Images*, 2016, pp. 178–193.

[17] A. Roy, B. Halder, U. Garain, and D. S. Doermann, "Machine-assisted authentication of paper currency: an experiment on Indian banknotes," *Int. J. Doc. Anal. Recognit.*, vol. 18, no. 3, pp. 271–285, 2015.

[18] 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 bank notes detection," in *2007 10th international conference on computer and information technology*, Dec. 2007, pp. 1–5, doi: 10.1109/ICCITECHN.2007.4579427.

[19] Ultra Electronics Forensic Technology Ltd. IBIS BULLETTRAX-3D, "The Only System to Automatically Image & Compare 3D images of Fired Bullets," 2009.

[20] A. M. Teredesai, M. A. Ahmad, J. Kanodia, and R. S. Gaborski, "CoMMA: a framework for integrated multimedia mining using multi-relational associations," *Knowl. Inf. Syst.*, vol. 10, no. 2, pp. 135–162, 2006.

[21] J. A. Rushing, H. S. Ranganath, T. H. Hinke, and S. J. Graves, "Using association rules as texture features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 845–858, Aug. 2001, doi: 10.1109/34.946988.

[22] Jiang Yun, Li Zhanhuai, Wang Yong, and Zhang Longbo, "Joining associative classifier for medical images," in *Fifth International Conference on Hybrid Intelligent Systems (HIS'05),* Nov. 2005, pp. 6 pp.-, doi: 10.1109/ICHIS.2005.67.

[23] R. Chaves, J. M. Górriz, J. Ram\'\irez, I. A. Illán, D. Salas-Gonzalez, and M. Gómez-R\'\io, "Efficient mining of association rules for the early diagnosis of Alzheimer's disease," *Phys. Med. Biol.*, vol. 56, no. 18, p. 6047, 2011.

[24] C. Ordonez and E. Omiecinski, "Discovering association rules based on image content," in *Proceedings IEEE Forum on Research and Technology Advances in Digital Libraries*, May 1999, pp. 38–49, doi: 10.1109/ADL.1999.777689.

[25] M. X. Ribeiro, A. J. M. Traina, C. Traina, and P. M. Azevedo-Marques, "An Association Rule-Based Method to Support Medical Image Diagnosis With Efficiency," *IEEE Trans. Multimed.*, vol. 10, no. 2, pp. 277–285, Feb. 2008, doi: 10.1109/TMM.2007.911837.

[26] D. Rohidin, N. A. Samsudin, and M. M. Deris, "Association rules of fuzzy soft set based classification for text classification problem," *J. King Saud Univ. - Comput. Inf. Sci.*, 2020, doi: https://doi.org/10.1016/j.jksuci.2020.03.014.

[27] Wenmin Li, Jiawei Han, and Jian Pei, "CMAR: accurate and efficient classification based on multiple class-association rules," in *Proceedings 2001 IEEE International Conference on Data Mining*, Nov. 2001, pp. 369–376, doi: 10.1109/ICDM.2001.989541.

[28] M.-L. Antonie, O. R. Zaiane, and A. Coman, "Associative classifiers for medical images," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2002, pp. 68–83.

[29] K. Song and K. Lee, "Predictability-based collective class association rule mining," *Expert Syst. Appl.*, vol. 79, pp. 1–7, 2017, doi: https://doi.org/10.1016/j.eswa.2017.02.024.

[30] M. Antonelli, P. Ducange, F. Marcelloni, and A. Segatori, "A novel associative classification model based on a fuzzy frequent pattern mining algorithm," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 2086–2097, 2015, doi: https://doi.org/10.1016/j.eswa.2014.09.021.

[31] L. Chen, "Curse of dimensionality," in *Encyclopedia of Database Systems*, 2009.

[32] A. K. Hmood and C. Y. Suen, "An ensemble of character features and fine-tuned convolutional neural network for spurious coin detection," *a chapter Lang. Process. Pattern Recognition, Intell. Syst.*, pp. 169–187, 2018.

[33] Jesse Robitaille, "'Canadian Coin News' Counterfeit coins cause for concern among Canadian collectors," 2015. .

[34] M. Zaharieva, R. Huber-Mörk, M. Nölle, and M. Kampel, "On Ancient Coin Classification.," in *VAST*, 2007, pp. 55–62.

[35] S. Modi, D. Bawa, and others, "Automated coin recognition system using ANN," *Int. J. Comput. Appl.*, vol. 26, no. 4, pp. 0975–8887, 2011.

[36] Y. Furuya, T. Ishida, I. Fukuda, and G. Yoshizawa, "Method and apparatus for sorting coins utilizing coin-derived signals containing different harmonic components," U.S. Patent 4971187, 1990.

[37] K. Okada, "Method and apparatus for discriminating coins or bank notes," U.S. Patent 4556140, 1985.

[38] M. Hida, T. Mitsui, and Y. Minami, "Forensic investigation of counterfeit coins," *Forensic Sci. Int.*, vol. 89, no. 1, pp. 21–26, 1997, doi: https://doi.org/10.1016/S0379-0738(97)00070-4.

[39] M. Hida, H. Sato, H. Sugawara, and T. Mitsui, "Classification of counterfeit coins using multivariate analysis with X-ray diffraction and X-ray fluorescence methods," *Forensic Sci. Int.*, vol. 115, no. 1, pp. 129–134, 2001, doi: https://doi.org/10.1016/S0379-0738(00)00309-1.

[40] S. Kim, S. H. Lee, and Y. M. Ro, "Image-based coin recognition using rotation-invariant region binary patterns based on gradient magnitudes," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 217–223, 2015, doi: https://doi.org/10.1016/j.jvcir.2015.08.011.

[41] A. Vadivelan and K. Praveen, "Implementation of coin recognition by extracting the local texture features with Gabor wavelet and LBP operator," in *Proc. Int. Conf. on Electrical Engineering and Computer Sciences (IEECS)(Coimbatore, India*, 2013, pp. 135–139.

[42] D. Pawade, P. Chaudhari, and H. Sonkambale, "Comparative study of different paper currency and coin currency recognition method," *Int. J. Comput. Appl.*, vol. 66, no. 23, pp. 26–31, 2013.

[43] A. Kavelar, S. Zambanini, and M. Kampel, "Word detection applied to images of ancient Roman coins," in *2012 18th International Conference on Virtual Systems and Multimedia*, 2012, pp. 577–580, doi: 10.1109/VSMM.2012.6365981.

[44] A. Kavelar, S. Zambanini, and M. Kampel, "Reading ancient coin legends: object recognition vs. OCR," *Proc. 37th Annu. Work. Austrian Assoc. Pattern Recognit.*, 2013.

[45] L. Shen, S. Jia, Z. Ji, and W.-S. Chen, "Extracting local texture features for image-based coin recognition," *IET Image Process.*, vol. 5, no. 5, pp. 394–401, 2011.

[46] M. Kampel, R. Huber-Mörk, and M. Zaharieva, "Image-Based Retrieval and Identification of Ancient Coins," *IEEE Intell. Syst.*, vol. 24, no. 2, pp. 26–34, Mar. 2009, doi: 10.1109/MIS.2009.29.

[47] C. Xu, "Research of coin recognition based on Bayesian network classifier," *Adv. Inf. Sci. Serv. Sci.*, vol. 4, no. 18, pp. 395–402, 2012.

[48] L. Shen, Sen Jia, Zhen Ji, and W. Chen, "Statictics of Gabor features for coin recognition," in *2009 IEEE International Workshop on Imaging Systems and Techniques*, May 2009, pp. 295–298, doi: 10.1109/IST.2009.5071653.

[49] L. J. Van Der Maaten and P. Poon, "Coin-o-matic: A fast system for reliable coin classification," in *Proc. of the Muscle CIS Coin Competition Workshop, Berlin, Germany*, 2006, pp. 7–18.

[50] M. Kampel and M. Zaharieva, "Recognizing ancient coins based on local features," in *International Symposium on Visual Computing*, 2008, pp. 11–22.

[51] R. Huber-Mörk, S. Zambanini, M. Zaharieva, and M. Kampel, "Identification of ancient coins based on fusion of shape and local features," *Mach. Vis. Appl.*, vol. 22, no. 6, pp. 983–994, 2011.

[52] N. Jain and N. Jain, "Coin Recognition Using Circular Hough Transform," *Int. J. Electron. Commun. Comput. Technol.*, vol. 2, no. 3, pp. 2249–7838, 2012.

[53] C. M. Velu, P. Vivekanandan, and others, "Indian coin recognition system of image segmentation by heuristic approach and houch transform (HT)," *Int. J. Open Probl. Compt. Math*, vol. 2, no. 2, pp. 224–271, 2009.

[54] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik, "Classification of coins using an eigenspace approach," *Pattern Recognit. Lett.*, vol. 26, no. 1, pp. 61–75, 2005, doi: https://doi.org/10.1016/j.patrec.2004.09.006.

[55] L. Van Der Maaten and E. Postma, "Towards automatic coin classification," in *In Proceedings of the EVA-Vienna*, 2006, pp. 19–26.

[56] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[57] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, 2010, doi: https://doi.org/10.1016/j.patcog.2009.08.017.

[58] M. Fukumi, S. Omatu, F. Takeda, and T. Kosaka, "Rotation-invariant neural pattern recognition system with application to coin recognition," *IEEE Trans. neural networks*, vol. 3, no. 2, pp. 272–279, 1992.

[59] R. Bremananth, B. Balaji, M. Sankari, and A. Chitra, "A New Approach to Coin Recognition using Neural Pattern Analysis," in *2005 Annual IEEE India Conference - Indicon*, Dec. 2005, pp. 366–370, doi: 10.1109/INDCON.2005.1590191.

[60] K. Hornik, M. Stinchcombe, H. White, and others, "Multilayer feedforward networks are universal approximators.," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[61] S. Marinai, M. Gori, and G. Soda, "Artificial neural networks for document analysis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 23–35, Jan. 2005, doi: 10.1109/TPAMI.2005.4.

[62] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[63] V. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag, Springer science & business media, 1995.

[64] L. Liu, Y. Lu, and C. Y. Suen, "Document image matching using probabilistic graphical models," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov. 2012, pp. 637–640.

[65] S. Todorovic and N. Ahuja, "Region-based hierarchical image matching," *Int. J. Comput. Vis.*,

vol. 78, no. 1, pp. 47–66, 2008.

[66] H. Bunke and K. Riesen, "Recent advances in graph-based pattern recognition with applications in document analysis," *Pattern Recognit.*, vol. 44, no. 5, pp. 1057–1067, 2011, doi: https://doi.org/10.1016/j.patcog.2010.11.015.

[67] M. Vento, "A long trip in the charming world of graphs for Pattern Recognition," *Pattern Recognit.*, vol. 48, no. 2, pp. 291–301, 2015, doi: https://doi.org/10.1016/j.patcog.2014.01.002.

[68] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, 2000.

[69] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005, doi: 10.1109/TPAMI.2005.188.

[70] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Recovery," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3, pp. 1099–1110, 2011, doi: 10.1109/TIFS.2011.2129512.

[71] X. Pan and S. Lyu, "Region Duplication Detection Using Image Feature Matching," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 4, pp. 857–867, Dec. 2010, doi: 10.1109/TIFS.2010.2078506.

[72] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002, doi: 10.1109/TPAMI.2002.1023800.

[73] W. Altaf, M. Shahbaz, and A. Guergachi, "Applications of association rule mining in health informatics: a survey," *Artif. Intell. Rev.*, vol. 47, no. 3, pp. 313–340, 2017.

[74] J. Deshmukh and U. Bhosle, "Image mining using association rule for medical image dataset," *Procedia Comput. Sci.*, vol. 85, pp. 117–124, 2016.

[75] A. J. T. Lee, R.-W. Hong, W.-M. Ko, W.-K. Tsao, and H.-H. Lin, "Mining spatial association rules in image databases," *Inf. Sci. (Ny).*, vol. 177, no. 7, pp. 1593–1608, 2007, doi: https://doi.org/10.1016/j.ins.2006.09.018.

[76] P. Rajendran and M. Madheswaran, "Novel fuzzy association rule image mining algorithm for medical Decision support system," *Int. J. Comput. Appl.*, vol. 1, no. 20, pp. 87–94, 2010.

[77] W. Hadi, G. Issa, and A. Ishtaiwi, "ACPRISM: Associative classification based on PRISM algorithm," *Inf. Sci. (Ny).*, vol. 417, pp. 287–300, 2017, doi: https://doi.org/10.1016/j.ins.2017.07.025.

[78] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge discovery in databases: An overview," *AI Mag.*, vol. 13, no. 3, p. 57, 1992.

[79] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in

databases," *AI Mag.*, vol. 17, no. 3, p. 37, 1996.

[80] B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri, "Mining numerical association rules via multi-objective genetic algorithms," *Inf. Sci. (Ny).*, vol. 233, pp. 15–24, 2013, doi: https://doi.org/10.1016/j.ins.2013.01.028.

[81] C. Zhang and S. Zhang, *Association rule mining: models and algorithms*. Springer-Verlag, 2002.

[82] M. Ceci and A. Appice, "Spatial associative classification: propositional vs structural approach," *J. Intell. Inf. Syst.*, vol. 27, no. 3, pp. 191–213, 2006.

[83] A. V. P. Muralidhar, "Tabu search based fast high utility item set mining," *J. Adv. Res. Dyn. Control Syst.*, vol. 9, no. 18, pp. 3152–3164, 2017.

[84] J. A. Delgado-Osuna, C. García-Martínez, J. Gómez-Barbadillo, and S. Ventura, "Heuristics for interesting class association rule mining a colorectal cancer database," *Inf. Process. Manag.*, vol. 57, no. 3, p. 102207, 2020, doi: https://doi.org/10.1016/j.ipm.2020.102207.

[85] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 1993, pp. 207–216.

[86] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," *Data Min. Knowl. Discov.*, vol. 8, no. 1, pp. 53–87, 2004.

[87] F. Guillet, P. Kuntz, and R. Lehn, "A genetic algorithm for visualizing networks of association rules," *Mult. Approaches to Intell. Syst.*, vol. 1611, pp. 145–154, 1999.

[88] M. J. del Jesus, J. A. Gamez, P. Gonzalez, and J. M. Puerta, "On the discovery of association rules by means of evolutionary algorithms," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 5, pp. 397–415, 2011.

[89] J. M. Luna, M. Pechenizkiy, M. J. del Jesus, and S. Ventura, "Mining Context-Aware Association Rules Using Grammar-Based Genetic Programming," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3030–3044, Nov. 2018, doi: 10.1109/TCYB.2017.2750919.

[90] J. M. Luna, J. R. Romero, and S. Ventura, "Grammar-based multi-objective algorithms for mining association rules," *Data Knowl. Eng.*, vol. 86, pp. 19–37, 2013, doi: https://doi.org/10.1016/j.datak.2013.01.002.

[91] E.-G. Talbi, *Metaheuristics: from design to implementation*, vol. 74. John Wiley & Sons, 2009.

[92] W. W. Cohen, "Fast Effective Rule Induction," in *Machine Learning Proceedings 1995*, A. Prieditis and S. Russell, Eds. San Francisco (CA): Morgan Kaufmann, 1995, pp. 115–123.

[93] J. A. Pedraza, C. Garc\'\ia-Mart\'\inez, A. Cano, and S. Ventura, "Classification rule mining with iterated greedy," in *International Conference on Hybrid Artificial Intelligence Systems*, 2014, pp. 585–596.

[94] J. R. Quinlan and R. M. Cameron-Jones, "FOIL: A midterm report," in *European conference on machine learning*, 1993, pp. 1–20.

[95] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," in *Proceedings of the 2003 SIAM International Conference on Data Mining*, 2003, pp. 331–335.

[96] M. Sheikhan and M. S. Rad, "Gravitational search algorithm--optimized neural misuse detector with selected features by fuzzy grids--based association rules mining," *Neural Comput. Appl.*, vol. 23, no. 7–8, pp. 2451–2463, 2013.

[97] M. Kaya and R. Alhajj, "A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining," in *The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ '03.*, May 2003, vol. 2, pp. 881–886 vol.2, doi: 10.1109/FUZZ.2003.1206547.

[98] M. J. Zaki, "Scalable algorithms for association mining," *IEEE Trans. Knowl. Data Eng.*, vol. 12, no. 3, pp. 372–390, May 2000, doi: 10.1109/69.846291.

[99] A. Telikani, A. H. Gandomi, A. Shahbahrami, and M. [Naderi Dehkordi], "Privacy-preserving in association rule mining using an improved discrete binary artificial bee colony," *Expert Syst. Appl.*, vol. 144, p. 113097, 2020, doi: https://doi.org/10.1016/j.eswa.2019.113097.

[100] Y. Djenouri, A. Bendjoudi, D. Djenouri, and M. Comuzzi, "GPU-based Bio-inspired Model for Solving Association Rules Mining Problem," in *2017 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, Mar. 2017, pp. 262–269, doi: 10.1109/PDP.2017.16.

[101] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965, doi: https://doi.org/10.1016/S0019-9958(65)90241-X.

[102] T.-P. Hong, Y.-F. Tung, S.-L. Wang, Y.-L. Wu, and M.-T. Wu, "A multi-level ant-colony mining algorithm for membership functions," *Inf. Sci. (Ny).*, vol. 182, no. 1, pp. 3–14, 2012, doi: https://doi.org/10.1016/j.ins.2010.12.019.

[103] C. Zou, H. Deng, J. Wan, Z. Wang, and P. Deng, "Mining and updating association rules based on fuzzy concept lattice," *Futur. Gener. Comput. Syst.*, vol. 82, pp. 698–706, 2018, doi: https://doi.org/10.1016/j.future.2017.11.018.

[104] C.-H. Chen, T.-P. Hong, and V. S. Tseng, "An improved approach to find membership functions and multiple minimum supports in fuzzy data mining," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 10016–10024, 2009, doi: https://doi.org/10.1016/j.eswa.2009.01.067.

[105] T. C.-K. Huang, "Discovery of fuzzy quantitative sequential patterns with multiple minimum supports and adjustable membership functions," *Inf. Sci. (Ny).*, vol. 222, pp. 126–146, 2013, doi: https://doi.org/10.1016/j.ins.2012.07.047.

[106] H. Zheng, J. He, G. Huang, Y. Zhang, and H. Wang, "Dynamic optimisation based fuzzy association rule mining method," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2187–2198,

2019.

[107] Z. Geng *et al.*, "A model-free Bayesian classifier," *Inf. Sci. (Ny).*, vol. 482, pp. 171–188, 2019, doi: https://doi.org/10.1016/j.ins.2019.01.026.

[108] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, "A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I," *IEEE Trans. Evol. Comput.*, vol. 18, no. 1, pp. 4–19, Feb. 2014, doi: 10.1109/TEVC.2013.2290086.

[109] V. Mangat and R. Vig, "Novel associative classifier based on dynamic adaptive PSO: Application to determining candidates for thoracic surgery," *Expert Syst. Appl.*, vol. 41, no. 18, pp. 8234–8244, 2014, doi: https://doi.org/10.1016/j.eswa.2014.06.046.

[110] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984.

[111] Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng, "Discovering fuzzy association rules using fuzzy partition methods," *Knowledge-Based Syst.*, vol. 16, no. 3, pp. 137–147, 2003, doi: https://doi.org/10.1016/S0950-7051(02)00079-5.

[112] C.-H. Chen, T.-P. Hong, Y.-C. Lee, and V. S. Tseng, "Finding active membership functions for genetic-fuzzy data mining," *Int. J. Inf. Technol. Decis. Mak.*, vol. 14, no. 06, pp. 1215–1242, 2015.

[113] M. Kaya, "Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules," *Soft Comput.*, vol. 10, no. 7, pp. 578–586, 2006.

[114] T.-P. Hong, C.-H. Chen, Y.-L. Wu, and Y.-C. Lee, "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions," *Soft Comput.*, vol. 10, no. 11, pp. 1091–1101, 2006.

[115] M. Sheikhan and M. Sharifi Rad, "Using particle swarm optimization in fuzzy association rules-based feature selection and fuzzy ARTMAP-based attack recognition," *Secur. Commun. Networks*, vol. 6, no. 7, pp. 797–811, 2013.

[116] C. Silverstein, S. Brin, and R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules," *Data Min. Knowl. Discov.*, vol. 2, no. 1, pp. 39–68, 1998.

[117] C.-M. Wang and Y.-F. Huang, "Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data," *Expert Syst. Appl.*, vol. 36, no. 3, Part 2, pp. 5900–5908, 2009, doi: https://doi.org/10.1016/j.eswa.2008.07.026.

[118] N. Kaoungku, K. Suksut, R. Chanklan, K. Kerdprasop, and N. Kerdprasop, "The Silhouette Width Criterion for Clustering and Association Mining to Select Image Features," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, 2018.

[119] X. Jin, A. Xu, R. Bie, and P. Guo, "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles," in *International Workshop on Data*

Mining for Biomedical Applications, 2006, pp. 106–115.

[120] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, no. Oct, pp. 1205–1224, 2004.

[121] Y. Saeys, P. Rouzé, and Y. de Peer, "In search of the small ones: improved prediction of short exons in vertebrates, plants, fungi and protists," *Bioinformatics*, vol. 23, no. 4, pp. 414–420, 2007, doi: 10.1093/bioinformatics/btl639.

[122] J. Xie, J. Wu, and Q. Qian, "Feature Selection Algorithm Based on Association Rules Mining Method," in *2009 Eighth IEEE/ACIS International Conference on Computer and Information Science*, Jun. 2009, pp. 357–362, doi: 10.1109/ICIS.2009.103.

[123] N. Kaoungku, K. Suksut, R. Chanklan, K. Kerdprasop, and N. Kerdprasop, "Data classification based on feature selection with association rule mining," in *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS2017)*, 2017, vol. 1, pp. 321–326.

[124] N. Kaoungku, K. Kerdprasop, and N. Kerdprasop, "A method to clustering the feature ranking on data classification using an ensemble feature selection," *Int. J. Futur. Comput. Commun.*, vol. 6, no. 3, pp. 81–85, 2017.

[125] P. N. Tan, M. Steinbach, and V. Kumar, "Introduction to data mining, addison wesley publishers," 2006.

[126] F. A. Thabtah, "A review of associative classification mining," *Knowl. Eng. Rev.*, vol. 22, no. 1, pp. 37–65, 2007.

[127] W.-C. Chen, C.-C. Hsu, and Y.-C. Chu, "Increasing the effectiveness of associative classification in terms of class imbalance by using a novel pruning algorithm," *Expert Syst. Appl.*, vol. 39, no. 17, pp. 12841–12850, 2012, doi: https://doi.org/10.1016/j.eswa.2012.05.009.

[128] W.-C. Chen, C.-C. Hsu, and J.-N. Hsu, "Adjusting and generalizing CBA algorithm to handling class imbalance," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5907–5919, 2012, doi: https://doi.org/10.1016/j.eswa.2011.11.113.

[129] F. Chen, Y. Wang, M. Li, H. Wu, and J. Tian, "Principal Association Mining: An efficient classification approach," *Knowledge-Based Syst.*, vol. 67, pp. 16–25, 2014, doi: https://doi.org/10.1016/j.knosys.2014.06.013.

[130] N. Abdelhamid, A. Ayesh, F. Thabtah, S. Ahmadi, and W. Hadi, "MAC: A multiclass associative classification algorithm," *J. Inf. Knowl. Manag.*, vol. 11, no. 02, p. 1250011, 2012.

[131] E. Baralis and P. Garza, "I-prune: Item selection for associative classification," *Int. J. Intell. Syst.*, vol. 27, no. 3, pp. 279–299, 2012.

[132] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Heart disease prediction using lazy associative classification," in *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, Mar. 2013, pp. 40–46, doi:

10.1109/iMac4s.2013.6526381.

[133] N. Abdelhamid and F. Thabtah, "Associative classification approaches: review and comparison," *J. Inf. Knowl. Manag.*, vol. 13, no. 03, p. 1450027, 2014.

[134] R. Bellman, R. Corporation, and K. M. R. Collection, *Dynamic Programming*. Princeton University Press, 1957.

[135] C. K. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. 16, no. 1, pp. 41–46, Jan. 1970, doi: 10.1109/TIT.1970.1054406.

[136] C. K. Chow, "An optimum character recognition system using decision functions," *IRE Trans. Electron. Comput.*, vol. EC-6, no. 4, pp. 247–254, Dec. 1957, doi: 10.1109/TEC.1957.5222035.

[137] C. De Stefano, C. Sansone, and M. Vento, "To reject or not to reject: that is the question-an answer in case of neural classifiers," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 30, no. 1, pp. 84–94, 2000.

[138] Y. Cui, Z. Geng, Q. Zhu, and Y. Han, "Review: Multi-objective optimization methods and application in energy saving," *Energy*, vol. 125, pp. 681–704, 2017, doi: https://doi.org/10.1016/j.energy.2017.02.174.

[139] R. J. Kuo, C. M. Chao, and Y. T. Chiu, "Application of particle swarm optimization to association rule mining," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 326–336, 2011, doi: https://doi.org/10.1016/j.asoc.2009.11.023.

[140] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Nov. 1995, vol. 4, pp. 1942–1948 vol.4, doi: 10.1109/ICNN.1995.488968.

[141] Y. Shi and R. C. Eberhart, "Parameter selection in particle swarm optimization," in *International conference on evolutionary programming*, 1998, pp. 591–600.

[142] C. Ou and W. Lin, "Comparison between PSO and GA for Parameters Optimization of PID Controller," in *2006 International Conference on Mechatronics and Automation*, Jun. 2006, pp. 2471–2475, doi: 10.1109/ICMA.2006.257739.

[143] S. Kung, "Applying genetic algorithm and weight item to association rule," Yuan Ze University, Taiwan, 2002.

[144] F. Chollet, *Deep Learning with Python*, First. Shelter Island, NY, US: Manning, 2018.

[145] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004, doi: https://doi.org/10.1016/j.imavis.2004.02.006.

[146] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing*, vol. 38, no. 1, pp. 99–112, 1994, doi: https://doi.org/10.1016/0165-1684(94)90059-0.

[147] R. C. Quinlan, "4.5: Programs for machine learning morgan kaufmann publishers inc," *San Fr. USA*, 1993.

[148] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Mach. Learn.*, vol. 11, no. 1, pp. 63–90, 1993.

[149] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," 1998.

[150] B. Liu, Y. Ma, and C.-K. Wong, "Classification using association rules: weaknesses and enhancements," in *Data mining for scientific and engineering applications*, Springer, 2001, pp. 591–605.

[151] J. Alcala-Fdez, R. Alcala, and F. Herrera, "A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems With Genetic Rule Selection and Lateral Tuning," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, Oct. 2011, doi: 10.1109/TFUZZ.2011.2147794.

[152] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.