CHAPTER 6
HANDWRITTEN CHARACTER RECOGNITION

Muhammad Arif Mohamad, Muhammad Aliif Ahmad, and Jamilah Mahmood

6.1 INTRODUCTION

The sensitivity of the chosen characteristics influences handwritten character recognition (HCR) accuracy. Numerous feature extraction techniques are now in the literature (Paul et al., 2010). The technique of producing feature information from raw data to increase sample variability between classes while decreasing sample variability within a class is known as feature extraction. For this reason, when representing each class, a set of qualities or characteristics must be found that may differentiate the traits of other classes while staying unaffected by character alterations (Zebari et al., 2020). The review of feature extraction techniques for character recognition was good (Kaur et al., 2023).

Chain code is a method for processing and analysing images based on the boundary feature representation. Freeman (1961) introduced the first chain coding method, Freeman chain coding (FCC). The approach to traverse the image and its starting point would significantly impact the chain-coding procedure's challenge. Even if the image were the same, the image characters' starting places would produce different FCC directions. Characters' starting points were opted arbitrarily. Therefore, the best answer was obtained. However, much investigation has not been done into chain code constructions via a single continuous

path with chain code length minimisation. A metaheuristic optimisation approach was used to solve this problem in extracting continuous chain code, which accurately reflects handwritten characters.

In summary, this study proposed a metaheuristic algorithm-based chain code feature extraction that uses the honey-badger algorithm (HBA). The FCC was made shorter by using HBA. The main problem of utilising FCC in representing character images is that the revisit walks, node branches, and commencement points affected how long the chain code was. HBA is used to create FCC, which could build the chain code that accurately depicts the image character to solve these problems.

6.2 CURRENT ISSUES AND CHALLENGES

There are three stages in HCR: Preprocessing, feature extraction and classification. The first stage, preprocessing, enhances the image character, which should be used for further processing. Preprocessing is almost one of the basic steps in HCR. Usually, it is used to remove noise and different variations in the data. It may include binarisation, noise reduction, data normalisation and data compression. Preprocessing aims to produce a clean character of handwritten character image that can be used directly and efficiently by the feature extraction stage.

The second stage, feature extraction, is a process to produce several characteristics or features from the image character. Feature extraction is related to the extraction method to find the most representative information, which minimises the within-class pattern variability while enhancing the between-class pattern variability (Li et al., 2017). Based on the literature review, two factors must be considered when selecting feature extraction methods: simple and efficient. Wang et al. (2018) stated that in finding the features, it should be noted that in order to avoid extra complexity and to increase the accuracy of recognition, more compact features are required. Furthermore, Gadekallu et al. (2021) stated that finding simple and efficient features for handwritten character recognition is still an active area of research.

Chain code by Freeman (1961), as one of the feature extraction methods under the geometrical and topological representation category, has been extensively applied for feature extraction purposes. This is due to its ease and minimal storage needs (Zalik et al., 2018). Furthermore, Dingli et al. (2018) stated that chain code is widely used to describe object borders in image processing, shape analysis and pattern recognition fields because of its simple and compact form of data representation and its suitability for fast processing. Subsequently, HCR literature has shown that chain code representation is still relevant in representing in HCR due to recent work on chain code for handwritten characters by (Dingli et al., 2018; Jangid & Srivastava, 2018; Naik & Desai, 2019).

Unfortunately, the problem with the chain code feature extraction process is that the chain coding process depends very much on how the image is traversed and the starting point of the traversing method. A start point of a character will produce a different chain code direction even though it is the same image. This means that the starting node of chain code construction influences its length. Moreover, the problem worsens when handwritten character recognition is involved since handwritten characters usually contain branches on each character. This causes difficulty in deciding which direction the traverse should continue, and a revisit to the previously visited node is often needed to visit all the nodes. Nasien et al. (2014), suggested that one continuous route is needed to solve such problems, which covers all the image nodes. Chain code construction using one continuous route has not been widely explored. Such a method would enable extracting and recognising such difficult characters and finding approximate solutions for chain code generation while minimising its length.

As the third stage of HCR, the classification stage intends to recognise the images of handwritten characters by using the extracted features to recognise the feature class based on the properties of the features. The success rate of HCR depends on the entire stage: preprocessing, feature extraction and classification stages. This study concentrates only on the feature extraction stage. So, this study focuses