BOUNDED BOX-ZONING INTEGRATED APPROACH FOR CHILDREN HANDWRITING RECOGNITION

NIK NUR ADLIN BINTI NIK QAUSBEE

UNIVERSITI SAINS MALAYSIA

2024

BOUNDED BOX-ZONING INTEGRATED APPROACH FOR CHILDREN HANDWRITING RECOGNITION

by

NIK NUR ADLIN BINTI NIK QAUSBEE

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

March 2024

ACKNOWLEDGEMENT

First and foremost, I extend my praises and thanks to Allah SWT, the Most Gracious and the Most Merciful, for His abundant blessings that guided me throughout my research study, enabling me to complete this journey. I am grateful for all the opportunities, trials, and strength bestowed upon me, contributing to the completion of this thesis. My humblest gratitude goes to the holy Prophet Muhammad (Peace be upon him) whose way of life has continuously guided me. I gained valuable experiences throughout the process, not only academically but also contributing to shaping my personality. I would like to express my deep and sincere gratitude to my research supervisor, Dr. Nur Intan Raihana Binti Ruhaiyem, for her guidance, understanding, patience, and most importantly for providing me with positive encouragement and warm spirit to complete this research study. Having her as my supervisor has been a great pleasure and honour. I would also like to express my gratitude to my co-supervisor, Dr. Low Hui Min, for her help and support throughout this research work. I am extremely grateful to my beloved parents for their unwavering love, prayers, support, care, and sacrifices in educating and preparing me for the future. This thesis would not have been possible without their consistent support. A very special gratitude is reserved to my dearest father, Nik Qausbee, my beloved mother, Suzana, and my brother Nik Shafiq Qadimi. I would like to offer my special thanks to all my beloved friends who were with me and supported me through thick and thin, especially thanking Sharifah Nurshahira and Muhammad Fawwaz Zahran. May God shower these mentioned personalities with success and honor in their lives.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

OCR	Optical Character Recognition
HCR	Handwriting Character Recognition
HMM	Hidden Markov Models
HOG	Histogram of Oriented Gradient
AI	Artificial Intelligence
ML	Machine Learning
DTC	Decision Tree Classifier
RFC	Random Forest Classifier
KNN	K-Nearest Neighbour
PCA	Principal Component Analysis
SVM	Support Vector Machine
NN	Neural Networks
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks
ICT	Information and Communication Technology
MHA	Minnesota Handwriting Assessment

KAEDAH INTEGRASI KOTAK SEMPADAN-ZON UNTUK PENGECAMAN TULISAN TANGAN KANAK-KANAK

ABSTRAK

Kajian ini memberi tumpuan kepada pengenalan set data aksara tulisan tangan kanak-kanak melalui integrasi pemprosesan imej dan pembelajaran mesin. Kotak sempadan, salah satu kaedah pengekstrakan ciri struktur terbaik, telah menunjukkan prestasi tinggi dalam saluran paip pengecaman aksara optik (OCR). Walau bagaimanapun, pelaksanaannya pada tulisan tangan kanak-kanak telah menunjukkan trend penurunan dalam pengesanan abjad. Begitu juga, pengezonan sebagai salah satu teknik yang berkuasa di bawah pengekstrakan ciri statistik, menunjukkan pengelasan yang baik, tetapi terhad dengan nilai ciri tanpa had dan tidak sesuai untuk aksara dengan variasi tinggi, seperti tulisan tangan kanak-kanak. Objektif kajian ini adalah untuk mengenal pasti abjad bahasa Inggeris yang signifikan berdasarkan ciri mereka, mencadangkan pendekatan bersepadu kotak sempadan dan pengezonan untuk meningkatkan saluran paip OCR, dan mengenal pasti ketepatan kaedah yang dicadangkan. Penilaian Tulisan Tangan Minnesota (MHA) telah digunakan untuk pengumpulan data, melibatkan sampel tulisan tangan dikumpulkan daripada 90 kanak-kanak berumur antara 6 hingga 9 tahun. Kajian ini kemudiannya melibatkan langkah pemprosesan imej, termasuk pengumpulan abjad ke dalam kumpulan 'kecil' dan 'ekor', pengekstrakan ciri menggunakan kaedah hibrid yang dicadangkan (iaitu kotak sempadan-pengezonan) dan pengelasan menggunakan kaedah rangkaian saraf input berbilang. Keputusan menunjukkan bahawa empat abjad yang signifikan untuk ‘kumpulan kecil’ adalah **a**, **c**, **e**, dan **o**, dan untuk ‘kumpulan ekor’ adalah **b**, **d**, **g**, dan **h**. Berbanding dengan kaedah sedia ada (kotak

sempadan), kaedah yang dicadangkan menunjukkan peningkatan dalam ketepatan untuk abjad yang signifikan ini: peningkatan hasil catatan a dari 16.1% kepada 87.1%, peningkatan hasil catatan c dari 77% kepada 80.5%, peningkatan hasil catatan e dari 65.4% kepada 76.5%, peningkatan hasil catatan o dari 82.5% kepada 84.1%, peningkatan hasil catatan b dari 28% kepada 68%, peningkatan hasil catatan d dari 73% kepada 82.5%, peningkatan hasil catatan g dari 43.8% kepada 46.9%, dan peningkatan hasil catatan h dari 69.2% kepada 73%. Sumbangan kajian ini termasuk pengumpulan dua set data menggunakan saiz fon yang berbeza (saiz 28 dan saiz 36), mencadangkan pengumpulan abjad menggunakan kotak sempadan tetap, mengenal pasti abjad yang signifikan dari setiap kumpulan, dan mencadangkan pendekatan hibrid yang menggabungkan kotak sempadan dan teknik pengezonan untuk pengekstrakan ciri untuk meningkatkan OCR. Sumbangan ini diharapkan dapat memberi manfaat untuk penyelidikan lanjut dalam menganalisis tulisan tangan kanak-kanak.

BOUNDED BOX-ZONING INTEGRATED APPROACH FOR CHILDREN HANDWRITING RECOGNITION

ABSTRACT

This study focuses on recognizing handwritten character datasets of children through the integration of image processing and machine learning. The bounding box, one of the best structural feature extraction methods, has demonstrated high performance in the optical character recognition (OCR) pipeline. However, its implementation in children's handwriting has shown a decreasing trend in alphabet detection. Similarly, zoning, a powerful technique under statistical feature extraction demonstrated good classification but is limited by having unlimited feature values and is not applicable for characters with high variations, such as children's handwriting. The objectives of this study are to identify significant English alphabets based on their features, propose a bounded box-zoning integrated approach to improve the OCR pipeline and identify the accuracy of the proposed method. The Minnesota Handwriting Assessment (MHA) was utilized for data collection, involving handwriting samples collected from 90 children aged between 6 to 9 years old. The study then proceeded with image processing steps, including alphabet grouping into 'small' and 'tail' groups, feature extraction using the proposed hybrid method (bounded box-zoning), and classification using the multi-input neural network method. The results show that four significant alphabets for 'small group' are **a**, **c**, **e**, and **o**, and for 'tail group' are **b**, **d**, **g**, and **h**. In comparison with the existing method (i.e., bounding box), the proposed method has shown improvements in accuracy for these significant alphabets: **a** recorded improvement from 16.1% to 87.1%, **c** recorded improvement from 77% to 80.5%, **e** recorded improvement from

65.4% to 76.5%, **o** recorded improvement from 82.5% to 84.1%, **b** recorded improvement from 28% to 68%, **d** recorded improvement from 73% to 82.5%, **g** recorded improvement from 43.8% to 46.9% and **h** recorded improvement from 69.2% to 73%. The contributions of this research include collecting two sets of data using different font sizes (size 28 and size 36), proposing alphabet grouping using a fixed bounding box, identifying significant alphabets from each group, and proposing a hybrid approach integrating bounding box and zoning techniques for feature extraction to improve OCR. These contributions are hoped to be beneficial for further research in analyzing children's handwriting.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Over the years, technology has revolutionized the human world and daily lives capable of supporting and streamlining numerous activities. Information technology has paved the way for technological advancement. In the last few decades, technology has progressed at a staggering rate. Optical Character Recognition, often abbreviated as OCR, is part of science and technology that has played an enormous role in text recognition software. OCR is the process of utilizing computer vision models to recognize and identify alphabets printed on paper or digital documents using electronic devices. Today, OCR not only aids in digitizing typewritten documents but also involves OCR models determining the shape of alphabets by detecting their patterns.

On the eve of the First World War in 1914, one of the earliest OCR systems was developed by Emanuel Goldberg with the technology that could read alphabets and convert them into telegraph code. During the early stage, OCR was considered an aid for blind people. However, it later evolved into a visiting field of research and development. One prominent area within this field is pattern recognition, which is concerned with processed pattern domains like characters, voices, and faces. OCR, in turn, is a subset of the pattern recognition field. The major function of pattern recognition is to categorize and classify data or objects based on the features extracted from known patterns. Its applications extend beyond trend analysis, aiding in making predictions. Pattern recognition finds diverse applications, such as in medical imaging and diagnosis for analyzing radiographic findings (Chung et al., 2019), biometric identification like speech recognition (Ahmed et al., 2021), and text document understanding in OCR.

Fundamentally, OCR can function in an offline mode, wherein alphabets are recognized after processing text documents. In this analogy, both Handwriting Character Recognition (HCR) and printed text documents are the outcomes of the offline OCR model. Alternatively, OCR can operate online, allowing alphabets to be recognized in real-time as writers or individuals are composing on digital devices. Nowadays, OCR offers a variety of applications and holds great potential. The ongoing need for research in character recognition is evident from the increasing variety of alphabets to be identified, emphasizing its continued relevance. Even today, OCR applications are widely used in areas such as writer identification and verification (Ghosh et al., 2019).

OCR plays a crucial role in enhancing the comprehension of handwritten text, especially in cursive or intricate handwriting, as it offers increased accessibility and efficiency. The inaccuracy of certain input data from various sources into a computer or system makes OCR the most crucial tool for enhancing data entry accuracy. Certain phases of OCR are applied in processing images before an image can be recognized. In this study, Chapter 2 and Chapter 3 delve into these phases with a comprehensive explanation of the primary component, which is the feature extraction phase.

1.2 Research Problem

OCR is one of the most successful methods in image processing for computerized handwriting recognition. It has successfully analyzed adult and adolescent handwriting across multiple global languages, including Arabic, Urdu, and

Mandarin. Previous research (Luo et al., 2022; Khan et al., 2019; Silva & Jayasundere, 2020) extensively explored the OCR model using datasets of adult handwriting. Based on the research by Bonneton-Botte et al. (2023), it is revealed that around 30% of typically developing children encounter challenges in handwriting formation. In developing the OCR model, each alphabet undergoes several phases of image processing, such as pre-processing, feature extraction, and classification. The feature extraction phase plays a vital role in OCR by extracting the appropriate features that yield the minimum classification errors. Specifically, when dealing with the diverse styles and varied aspect ratios of children's handwriting, proper feature extraction within the OCR pipeline becomes imperative.

One of the best and most popular techniques of structural feature extraction is the bounding box. It specifies the position of the alphabet in terms of its center, width, and height, providing a clear indication of the alphabet's location. After identifying each character, they are individually cropped. Each alphabet is encompassed by its bounding box, effectively separating one from another. A study conducted by Iza and her research team, which implemented the bounding box technique to extract the features of the alphabet, found its effectiveness is limited due to the lack of features employed in this method. The bounding box technique relies on a single feature, specifically the measurement of the area only. Also, there is room for improvement when implementing this technique in children's datasets (Iza Sazanita, et al., 2019).

On the other hand, in statistical feature extraction, zoning stands out as an effective method for handwriting analysis. Herwanto et al., (2019) presented a new approach to the zoning method by computing the number of black pixels in each zone and storing this information in an array. The study's findings showed that the accuracy is influenced by an increase in the number of added zones. This demonstrates that increasing the number of zones can offer more precise information to the classification algorithm for recognizing the alphabet. However, it is important to note that this technique provides unlimited feature values and is not recommended for a character with high variation (Sarma et al., 2019). Despite these weaknesses, along with the acknowledged benefits and promising outcomes, it is believed that utilizing both techniques can improve the feature extraction of the OCR pipeline.

1.3 Research Question

I. How significant alphabets can be identified from the children handwriting?

II. How feature extraction can be improved using bounding box method?

1.4 Research Objective

I. To group small and tail alphabets for significant alphabets identification.

II. To propose an improved feature extraction method by integrating bounding box and zoning methods.

1.5 Research Scope

The scope of this study is restricted to the selection of image-processing techniques with the help of a multi-input neural network model. The empirical study in this research is confined to 90 children aged six to nine years. A dataset containing handwriting samples from these children, totaling 6120 alphabets from two different font sizes: 28 and 36.

1.6 Research Contribution

The main contribution of this research is proposing a hybrid bounding box and zoning technique to improve feature extraction in the OCR pipeline. The sub-contributions from this research are as follows:

I. Propose alphabet grouping using a fixed bounding box and identify the significant alphabets from each group.

II. Propose an improved feature extraction method namely bounded box-zoning.

1.7 Thesis Outline

The research is divided into five chapters and references. **Chapter One** contains the introduction to the research; a general background, the research problems, questions, objectives, scope, and contributions. **Chapter Two** reviews the existing literature, focusing on the fundamental phase of OCR and its implementation in children's handwriting. **Chapter Three** discusses the research methodology, which includes a detailed description of each phase. **Chapter Four** presents the results and discussions while **Chapter Five** concludes the study and provides recommendations for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The origins of OCR can be traced back to the successful creation by the Russian scientist Tyurin in 1900, and its foundation in the early 1870s as an aid to the visually handicapped. In the mid-1950s, the modern computerized version of OCR emerged with a flatbed scanner capable of high-speed scanning. These advancements significantly accelerated character recognition, reduced cost, and facilitated the scanning of a wide range of text and documents (Govindan & Shivaprasad, 1990). Since then, extensive studies have been conducted, and numerous research papers in the field of character recognition have been published by various researchers, focusing on the recognition of adult and adolescent levels for Arabic, Mandarin, and Urdu characters. On the other hand, relatively little research has been conducted or is still ongoing for children's handwriting, primarily due to the complexity of children's handwriting.

2.2 Fundamental Phases of Optical Character Recognition

OCR systems can be categorized as Handwriting Character Recognition (HCR) and machine-printed character recognition (Bhasali & Kumar, 2013). HCR is the process of converting scanned manuscripts into editable and searchable text documents. This system can be further divided into two types: offline and online character recognition. In offline character recognition, the writing is typically captured by a scanner, and the completed writing is available as an image. Meanwhile, online character recognition acquires live handwriting for recognition. The analysis of handwriting in HCR generally involves several phases. The selection of techniques to be implemented is determined by the character input.

2.2.1 Data Acquisition

An image must be captured by a camera and transformed into a controllable entity before any video or image processing begins. This process is commonly abbreviated as image acquisition. It involves the digitizing of an optical image which measures real-world physical phenomena, into an array of numerical data that can be manipulated by a computer and software (Kumar & Chitra, 2020). Hence, a computer and software can analyze image data and present solutions. Performing image acquisition is always the first and most important step in the workflow sequence since it provides input data for the whole process. The image must be in a specific format like Joint Photographic Experts Group (JPG), Portable Network Graphics (PNG), and Bitmap (BMP). JPG image is used when working with more complex or opaque images, and photography albums (Sarma et al., 2019). PNG images are essential when working with transparent images and those featuring hard lines (Varma et al., 2020). Meanwhile, BMP images are very useful when dealing with images that will undergo alterations and include color gradations (Dome & Sathe, 2021).

A research study conducted by Thi and Takato has developed a handwritten database for recognizing English handwritten characters collected from children with diverse backgrounds. The authors collected all the input alphabets during the data acquisition phase (Thi & Takato, 2020). Meanwhile, Zin and his research team worked with a self-collected handwritten dataset. The term ‘self-collected’ refers to the authors creating a simple application to collect handwritten data on an Android tablet, with the features of a screen on which to draw, and two buttons; save and clear.

All data was gathered from children ranging from kindergarten to primary school. In addition, the authors mentioned that this method of data collection facilitates easy access to diverse data from various countries, as it relies on self-collected handwritten data using an Android tablet (Zin et al., 2021).

2.2.2 Pre-preprocessing

The acquired and collected data usually comes from various sources, often being messy and unorganized. It is imperative to organize, clean, and standardize the data before feeding it into the model system. The pre-processing phase is responsible for cleaning up and enhancing the quality of the image by objectively implementing image processing techniques to obtain a more concise representation of the image (Awel & Abidi, 2019). This phase is crucial to ensure the proper functioning of models or programs, delivering the expected output. Converting the image to a suitable format in this phase enhances the effectiveness and ease of processing in the next phases (Nanda & Goswami, 2021). The success rate of the pre-processing phase directly influences the recognition rate of the application. Several empirical studies have focused on the pre-processing phase and its techniques, highlighting their common use in analyzing the handwriting of children.

When capturing an image, the scanner records the object's image, including its color. Color images typically consist of three different layer images, a red scale image, a green scale image, and a blue scale image. Storing a single-color pixel in a Red, Green, or Blue (RGB) color image requires 24 bits (8 bits for each color component) instead of the 8 bits needed for a single greyscale image. Indeed, color may provide limited benefits in many applications. Thus, a greyscale technique is required to transform the colored image into a greyscale-intensity image (Herwanto et al., 2019;

Duth & Amulya, 2020; Imran et al., 2020). Greyscale images are considerably more convenient for a wide range of tasks. In a study, Iza and her research team applied the greyscale conversion process to handwritten colored images, which are more challenging to process compared to greyscale images (Iza Sazanita et al., 2019).

Generally, all documents are saved in computer memory in greyscale, utilizing a maximum of 256 different grey values ranging from 0 to 255 (Lamsaf et al., 2018). Document images must be processed multiple times to obtain the necessary data. The most efficient approach is to use binary images, reducing the time required for image extraction. The binarization technique, often referred to as threshold, categorizes images based on their grey values (0 to 255) or color into two classes of pixels (0 and 1) (Imran et al., 2020; Murugan et al., 2020).

The binarization technique is responsible for segregating the foreground layer, containing alphabet information, from the background layer, which may have noise. Binarized images of high quality tend to be more accurate in character recognition due to less noise present in original images. Binary images can be obtained through two types of thresholding; adaptive (local or variable) threshold and global threshold (Herwanto et al.., 2019; Duth & Amulya, 2020; Dome & Sathe, 2021). In the research by Majumder, et al., (2019), the thresholding method is clearly explained, allowing the alphabets to stand out in the images.

Local thresholding is commonly used when processing images with unstable intensity levels or poor-quality images. In contrast, global thresholding is a straightforward operation that extracts objects from the background by comparing image values with a threshold value (Davies, 2018). (Dome & Sathe, 2021) presented a global grey-level threshold histogram using the Otsu method. This method will exhaustively search for the threshold that minimizes the variance within the class. The authors mentioned that this process enhances the processing speed and reduces storage space. Binarization is the crucial landmark in the pre-processing phase (Zin et al., 2021; Thi & Takato, 2020).

The pre-processing phase is crucial in enhancing image detail by eliminating noise. (Sarma et al., 2019; Imran et al., 2020). It addresses issues such as salt and pepper, like dots in an image, uneven contrast, interfering strokes, etc., which do not contribute any significance to the output. Noise may occur throughout the image acquisition process due to reproduction and transmission. Various researchers have outlined specific methods to eliminate noise based on their data. (Dome & Sathe, 2021) presented methods such as Gaussian blur, mean, and bilateral filter to filter the noise and enhance the quality of the image text. Even when the image has minimal noise, its removal is essential for improved accuracy. Thi and Takato implemented a noise elimination technique in their work (Thi & Takato, 2020). The most common technique used for noise removal is the median filter, which effectively removes unwanted noise and distortion while preserving the edges of the alphabet to enhance some image features that are important for further processing (Duth & Amulya, 2020).

Besides, the Gaussian blur filter is used as a preprocessing technique to reduce noise with high-frequency components and remove speckles within the image. It is needed to eliminate isolated pixel noise. A study by (H Parikshith et al., 2021) proved that the Gaussian blur filter is a better choice to remove pixel noise in the alphabet image. It can effectively mute various types of noise present in digital images captured under conditions such as sensor noise induced by poor lighting, excessive temperature, or transmission issues (Senthil Kumar et al., 2021). The process of the

Gaussian blur filter starts with independently blurring each pixel in the desired area of the image, cutting out noise with higher frequencies. This linear filter is mostly used as a simple method to effectively blur the edges of the alphabet. To sum up, the Gaussian blur filter is an appropriate technique for detailed noise removal.

The Gaussian blur filter is employed to reduce the number of noise speckles, causing the shapes of the alphabet to be blurred. Hence, the morphological filter is needed to preserve the sharp edges of the alphabet in the image. In essence, the morphological filter aims to improve the sharpness of the images. In various studies, it is used to help noise removal and to smooth the alphabet's body. The dilation algorithm, known as the pseudo-inverse of erosion, is used to expand the alphabet through vector addition. It is important to note that erosion is not the opposite of dilation; instead, it functions as the reverse, used to shrink the alphabet (Sarma et al., 2019).

Another morphological operation involves opening and closing, which can be derived by combining the basic operations and using the same structural element. The opening operation is equivalent to erosion and aids in smoothing contours while eliminating small objects in the alphabet's image. The closing operations, resembling dilation, tend to narrow smooth sections of contours, fuse narrow breaks, eliminate small holes in the image's contour, and fill small gaps in the image (Dome & Sathe, 2021). Researchers (Thi & Takato, 2020) have applied morphological operations to correct missing alphabets and remove noise.

Below is an overview of the main techniques used in the pre-processing phase to analyze the children's handwriting data, based on previous research studies. It starts with greyscale and binarization encompasses several methods that can be selected based on specific data requirements. Following greyscale and binarizations, multiple noise removal techniques, including methods such as Gaussian blur filter and morphological filter are highlighted in this part. A concise view of some key techniques considered in this phase is presented in Table 2.1.

Table 2.1 The pre-processing techniques for character recognition

Pre-processing	**Author**	**Objective**	**Method**	**Advantage**
Greyscale	▪ (Herwanto et al, 2019) ▪ (Duth & Amulya, 2020) ▪ (Imran et al., 2020)	▪ To compress the color intensity ▪ It helps to simplify the algorithm. ▪ It eliminates the complexities	▪ Lightness ▪ Average ▪ Luminosity	▪ Simplify the algorithm. ▪ Reduce the data to threefold
Binarization	▪ (Sarma et al., 2019) ▪ (Herwanto et al, 2019) ▪ (Silva & Jayasundere, 2020)	▪ To separate the image into foreground text and background ▪ To convert a greyscale image into a black-and-white image ▪ To enhance the visual quality of an image	▪ Fixed thresholding ▪ Otsu method ▪ Adaptive method	▪ Decrease computational load. ▪ Increase the efficiency of the systems. ▪ Allow easy separation of an object from the background
Gaussian Blur	▪ (Sarma et al.,	▪ To	▪ Mean filter	▪ Very effective for

Filter	2019) ▪ (Herwanto et al, 2019) ▪ (Imran et al., 2020)	reduce high-frequency noise ▪ To reduce speckles		filtering Gaussian noise ▪ Remove detail noise. ▪ Computationally efficient
Morphological Filter	▪ (Sarma et al., 2019) ▪ (Dome & Sathe, 2021) ▪ (Thi & Takato, 2020)	▪ To remove the imperfections in the structure of the image ▪ To sharpen and smoothen the binary image	▪ Dilation ▪ Erosion ▪ Opening ▪ Closing	▪ Can detect the boundary (edge) of the object effectively. ▪ Quite effective for smoothing binary image

2.2.3 Feature Extraction

When dealing with extensive and seemingly redundant data, with a lot of data but minimal information for an algorithm to process, it needs to be transformed into a smaller representation set of features known as a feature vector. The extracted features are merged into a vector, which is then utilized to recognize the input and target output units. This process of converting the input data into a set of features is known as feature extraction. The primary purpose is to calculate the attributes of each alphabet while retaining as much information as possible from a large amount of data, thereby improving the speed and accuracy of the classifier in the recognition system (Varma et al., 2020). At this point, feature extraction techniques are used to identify a set of features that define the characteristics of the underlying alphabet as precisely and uniquely as possible.

These feature extraction techniques can be classified into statistical and structural feature extraction (Mustafa & Alia, 2019) as shown in Figure 2.1.

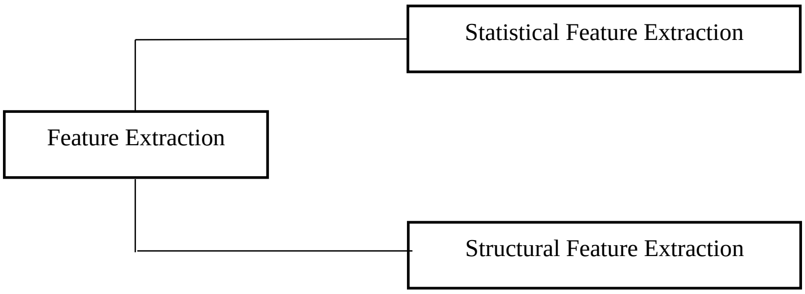

Figure 2.1 Type of feature extraction techniques.

Currently, statistical-based feature extraction techniques are widely used applications in the fields of knowledge discovery and machine learning. Based on the previous research, most of the ML applications rely on statistical-based computations to extract relevant features from the processes. The statistical-based feature extraction is accomplished by arranging points that constitute the alphabet matrix (Soora & Deshpande, 2017). These features provide high speed, low complexity, and the ability to handle certain style variations. There are various methods employed to implement statistical feature extraction**.**

The zoning feature extraction method is one of the most popular and efficient techniques in implementing statistical techniques in character recognition systems (Huang, et al., 2021). This technique handles variability and differences in alphabet patterns, including both cursive and printed alphabets. In essence, this technique extracts features by counting the number of black pixels in each alphabet zone. The process includes dividing the alphabet image into several smaller sections called zones. In each zone, the density of the alphabet is extracted, which reflects the ratio of the black pixels forming the alphabet to the entire area of the zone.

The zoning technique is typically applied to cover the entire part of the image as shown in Figure 2.2. This technique's advantage lies in its robustness to small variations, facilitating a smooth implementation process. Applying a pixel-based algorithm with the zoning technique may be subsidiary to achieving a good recognition rate even without considering specific pre-processing techniques. Additionally, the zoning-based feature extraction technique proves suitable for real-shaped characters. In a study by Herwanto et al. (2019), a novel approach to the zoning method was introduced by computing the number of black pixels in each zone and storing this information in an array. The result of the study indicated that the accuracy is affected by an increase in the number of added zones. This demonstrates that a higher number of zones offers more precise information to the classification algorithm for alphabet recognition. However, this technique is not recommended for characters with high variation due to its unlimited feature values (Sarma et al., 2019).

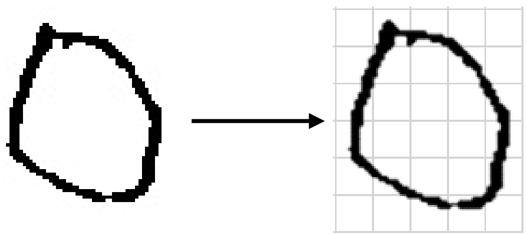

Figure 2.2 An example of zoning method.

On the other hand, structural-based feature extraction is one of the popular techniques that detects features via global and local characteristics of the alphabet. Some of these features have high tolerances for distortions and variances in style, including stroke directions, number of endpoints, number of holes, ratio, cross points, loops, etc. Structural feature techniques can be grouped into two, which are extracting and counting topological structures, and measuring and approximating geometrical properties (Soora & Deshpande, 2017; Mohamad, et al., 2016).

For extracting and counting topological structures, the features of the alphabet such as lines, curves, splines, extreme points, maxima and minima, cups above and below a threshold, openings, isolated dots, a bend between two points, horizontal curves at top or bottom, straight strokes between two points, ascending, descending and middle strokes as well as relations among the strokes, are considered as a vector (Mohamad, et al., 2016). In the method of measuring and approximating the geometrical properties, the alphabets are described by measuring geometric parameters such as the ratio between the width and height of the bounding box of a character, the relative distance between the last point and the last y-min, the relative horizontal and vertical distances between first and last points, the distance between two points of character, comparison of lengths between two strokes, the stroke width, upper and lower masses of words, word length curvature or change in curvature.

Structural techniques are widely used in the feature extraction step of optical character recognition due to their advantages. These techniques are versatile, making them applicable for multilingual languages, while also reducing computational costs and enhancing accuracy and efficiency in character recognition for identification and verification purposes. In summary, the structural feature extraction techniques are potentially enough to handle small variations in characters effectively.

Several studies have highlighted the significance of the bounding box technique as it is a very popular and extensive method among various structural feature extraction techniques. The bounding box is one of the most popular techniques in the feature extraction phase (Pham, et al., 2020). The bounding box consists of annotation markers drawn around the alphabet in an image. As the name implies, the boxes are rectangular. This technique applies a rectangular frame to outline the alphabet, creating a box that surrounds the alphabet in the image (Duth & Amulya, 2020).

The bounding box technique specifies the position of the alphabet using the coordinates of its top-left and bottom-right points. It further defines the alphabet concerning its center, width, and height, allowing a comprehensive representation of its location. Following the identification of each character, they are individually cropped, with each alphabet covered by a separate bounding box, thereby separating them from one another. A study by (Iza Sazanita, et al., 2019), which implemented the bounding box technique to extract the features of the alphabet, found that it is not within an optimal range and requires improvement, especially when applied to children's datasets. Tan’s research team also observed a consistent reduction in errors in the detected alphabets when using a bounding box (Tan, et al., 2022).

Many of the proposed techniques for extracting handwriting features in a specific language may be successful, especially when applied to text written in the standard form by normal adults. However, if the sample data is written in a style that deviates from the standard alphabet, such as when written by a child or a person with a medical condition, certain feature extraction techniques may not be effective. This limitation arises since it heavily depends on specific characteristics of the handwriting alphabet. In this context, Table 2.2 presents a review of feature extraction techniques, distinguishing between statistical and structural-based approaches.

Table 2.2 The feature extraction techniques for handwriting.

Feature Extraction	Method	Author	Data Set	Age Group	Accuracy
Statistical	**Zoning**	(Raj, Abirami, & Shyni, 2023)	HP dataset	Multi-age	90.31%
			Tamil data (self-collection)	Multi-age	87.26%
			Tamil dataset - Thirukkural	Multi-age	89.83%
		(Deshmukh & Kolhe, 2023)	Modi characters	Multi-age	83.09%
		(Prasad & Dubey, 2022)	English characters (self-collection)	Multi-age	92.6%
		(Abhale, et al., 2021)	English alphabets (self-collection)	Multi-age	80%
		(Huang, et al., 2021)	Pashto characters (self-collection)	Multi-age	76.42%
Structural	**Bounding Box**	(Luo, et al., 2022)	English vocabulary	Adult	55.25%
		(Silva & Jayasundere, 2020)	Sinhala characters (self-collection)	Adult	66%
		(Iza Sazanita, et al., 2019)	English alphabets	Children	57.50%
		(Ingle, et al., 2019)	English Vocabulary	Adult	59.6%

2.2.4 Classification

Various alphabet patterns emerge in handwriting, demanding a classification technique to distinguish between different groups or classes of alphabets. Classification is one of the most critical phases in the recognition model, since it needs to distribute unpredicted input data to their corresponding class, creating groups with homogeneous qualities, and segregating different inputs into different classes (Huang et al., 2021; Imran et al., 2020; Murugan et al., 2020; Herwanto et al., 2019).

In simple words, classification aids in isolating the feature space into different classes based on the predefined decision rule.

Before comparing the feature vectors corresponding to the input alphabet, the classifier should be trained on several patterns. When selecting the best classifier for a particular study, several factors need consideration, including the number of free parameters, available training set, and so on. Conversely, the effectiveness of a classifier depends on its parameters and technique (Verma, 2021; Surinta, 2016). To achieve optimal results in the recognition and classification model, various techniques are available, all ultimately grounded in image processing and neural networks, as shown in Figure 2.3.

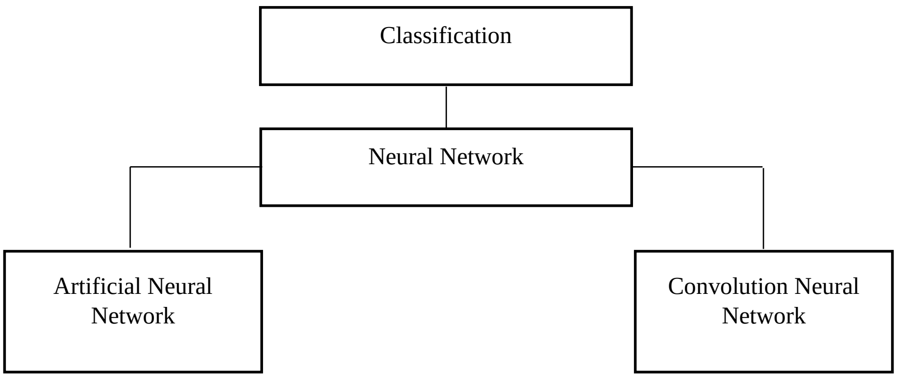

Figure 2.3 Types of classification in neural network.

The process of classification can be performed using either of the two techniques: classical techniques or soft computing techniques. Classical techniques are primarily based on image processing, for example, structural technique, statistical technique, and template matching (Xu, et al., 2021). As Artificial Intelligence (AI) becomes integrated into Machine Learning (ML), it increasingly spans various research domains, consistently achieving very good results. In the field of handwriting character recognition, ML has implemented various methods such as Logistic

Regression, Support Vector Machines, and Neural Networks. These techniques are fundamentally grounded in ML (Hamad & Kaya, 2016).

Within the ML domain, two different approaches are available: supervised or unsupervised learning. Supervised learning uses the ML approach to analyze labeled datasets, aiding in the training or supervising models for classifying data or predicting output for unforeseen data. Supervised learning models serve two main tasks: classification, and regression. Classification precisely classifies test data into specific groups, while regression uses algorithms to understand the relationship between dependent and independent variables (Khan et al., 2021). Common models or algorithms in supervised learning include neural networks.

The field of alphabet classification is defined by heuristic rationale, as people can recognize and record alphabets through training and experience. Thus, neural networks, being heuristic, are ideally suited for this type of issue. Neural Networks (NN), specifically a subfield of ML and the backbone of deep learning algorithms, are information processing paradigms or computer systems inspired by the biological NN that make up the human brain (Aqab & Tariq, 2020). Generally, NN is based on two important types: Artificial Neural Network (ANN) and Convolutional Neural Network (CNN).

Several researchers have discovered and proposed the use of NN in developing character recognition systems or applications. NN has proven to be an efficient approach for recognizing handwriting, as supported by various pieces of evidence. (Aqab & Tariq, 2020) presented a study, focusing on offline cursive handwritten character recognition, mainly to recognize the student’s and lecturer’s handwritten notes. This study utilized a feed-forward backpropagation neural network along with a new feature extraction approach, namely, diagonal feature extraction. The experimental results showed the highest recognition accuracy of 97% is obtained through the 54-feature diagonal approach. This means that ANN has a higher rate of recognition.

A research study conducted by (Nashif, et al., 2018) proposed work on three subfields of offline and online handwriting recognition: English alphabets, numeric characters, and individual signatures using multilayer CNN with an administered back propagation-learning algorithm. The accuracy rate of the proposed model met the research expectations, achieving a range of 97%. Meanwhile, (Abdul Hamid & Amir Sjarif, 2017) presented a study on interpreting intelligible handwritten input using three classifications: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and CNN. The authors concluded that CNN is more suitable for learning non-linear models based on the achieved results. However, it is significant to emphasize that CNN has a drawback: it takes a long time for training processes.

(Isthiaq & Saif, 2020) developed an Optical Character Recognition (OCR) system specifically designed for printing Bangla alphabets. In this approach, local and global features are extracted from the images, and the ANN classifier with backpropagation and one hidden layer is utilized for the classification task. The authors used their dataset of 297,898 Bangla single alphabet images of different fonts to train the ANN, as it requires big data for the training process. The result shows an accuracy of 69.82%. At the end of the experiment, the authors found that ANN not only requires a huge dataset for training but also needs consideration of other elements, such as the pre-processing phase. The author also mentioned the need for additional pre-processing techniques to increase the accuracy.

Table 2.3 Neural Network techniques in classification and recognition for handwriting.

Classification and Recognition	**Author**	**Data Set**	**Age Group**	**Accuracy**
ANN	(Rahman, et al., 2019)	Bengali alphabets (self-collection)	Multi-age	74.47%
ANN	(Khan, et al., 2019)	Pashto alphabets (self-collection)	Adult	72%
ANN	(Aqab & Tariq, 2020)	English alphabets (self-collection)	Adult	97%
ANN	(Isthiaq & Saif, 2020)	Bangla alphabets (self-collection)	Adult	69.82%.
CNN	(Nashif, et al., 2018)	English alphabets, numeric characters, and individual signature (self-collection)	Adult	97%
CNN	(Abdul Hamid & Amir Sjarif, 2017)	MNIST Dataset	Multi-age	99.26%
CNN	(Ullah & Jamjoom, 2022)	Arabic handwritten characters dataset	Adult	96.78%
CNN	(Chilvery, et al., 2022)	IAM dataset	Multi-age	93%

2.3 The Implementation of OCR in Children's Handwriting

Most models for handwriting recognition and OCR focus on adult and adolescent handwriting, as well as digit recognition, in comparison to children's handwriting. Nowadays, the ability to recognize children's handwriting has become crucial for various reasons, including educational and health-related purposes. In this aspect, the early years of children in elementary education are both challenging and important in developing effective and quality education for a lifetime. Their writing skills, especially, need improvement through encouragement. To achieve this, the handwritten characters and numerals done by children, particularly those living in less developed nations, must be accurately recognized to allow the establishment of remedies and solutions for improvements.

To the best of our knowledge, only a limited number of previous works have focused on OCR in children’s handwriting. In a recent study (Najwa & Isra, 2021), a handwritten database was developed for Arabic handwriting recognition systems. The authors developed a model for offline Arabic handwriting recognition using deep neural networks. This offline handwriting recognition model has been tested on Arabic-speaking school children aged between the ages of 7 and 12 in Riyadh, Saudi Arabia. To improve handwriting recognition performance, the authors proposed the use of CNN, as its effectiveness has been demonstrated in various applications including image segmentation, object detection, and image classification. The achieved recognition rate was 88%.

In a paper by (Alrobah & Albahli, 2021), a similar hybrid deep model was used in recognizing Arabic handwriting characters. The authors improved their approach by developing a hybrid model, combining CNN with SVM and eXtreme

Gradient Boosting (XGBoost) classifiers to recognize Arabic handwriting by children. CNN is utilized for the feature extraction of Arabic character images, which are then passed on to Machine Learning classifiers. The recognition rate for Arabic characters reached 96.3% for 29 characters.

A research study conducted by Corbille and his team has investigated both early and late fusions of multiple channels using different CNNs for online children's handwriting recognition. In the pre-processing phase, the authors chose to use linear normalization, spatial sampling, orientation alphabets, and direction alphabets. Multiple channels were used to represent dynamic information alongside static images, which represent the online handwriting, as input for CNN using the Latin handwriting dataset. The recognition rates for early fusion across three network architectures were 91.93% for LeNet-5, 94.33% for ResNet-18, and 94.84% for VGG-11. For late fusions, the recognition rates decreased for LeNet-5 (88.15%), ResNet-18 (93.19%), and VGG-11 (93.77%) (Corbille et al., 2020).

Santiari and Rahayuda presented an online handwriting character recognition system to analyze children's handwriting, using the K-NN algorithm. Children need to write capital letters A to H based on the sample given on a touchscreen tablet. Pre-processing is performed using noise removal technique, angle correction, and resizing. Statistical features are then extracted from the text image and fed into the recognizer. Considering the K-NN algorithm, the authors mentioned that it needs a large amount of memory and is slow in classification. The authors opted for the KNN algorithm considering it as the simplest machine learning algorithm. However, the achieved recognition rate was only 78.75% (Santiari & Rahayuda, 2022).