Handwritten Mathematical Symbol Recognition Using Machine Learning Techniques: Review

Syeda Aliya Firdaus(✉) and K. Vaidehi

Department of Computer Science and Engineering,
Stanley College of Engineering and Technology for Women,
Abids, Hyderabad 500 001, India
syedaaliya2407@gmail.com, kvaidehi@stanley.edu.in

Abstract. Handwritten character/symbol recognition is an important area of research in the present digital world. The problems such as recognition of handwritten characters/symbols, which may be written in different styles when it is recognized can makes job of the human easier. Mathematical expression recognition using machines has become a subject of serious research. The main motivation for this review work is both recognizing of the handwritten mathematical symbol, digits and characters which will be used for mathematical expression recognition.

Keywords: Mathematical symbol recognition · Character segmentation · Character recognition · Mathematical expression recognition

1 Introduction

Handwriting recognition (HWR) [1] is a computer process performed to obtain and understand the handwritten input such as touch-screen, paper documents, photographs and other devices. The images of the written text papers is called as "off line" taken by optical image scanning (or intelligent word recognition). The motion of the pen tip felt generally on a pen-based computer screen surface can be called as "on line", a easier task as there are more options available.

Handwriting recognition primarily follows the process of optical character recognition (OCR). The handwriting recognition system handles and includes formatting of the document, performs the correct segmentation into characters and also finds the most possible words.

OCR [2] can be both mechanical or electronic converter. The conversion includes conversion of handwritten image, typed image or printed text into machine-encoded text, taken from a photograph of a document, a scanned document, a scene-photo from subtitle text superimposed on an image. OCR is usually an "offline" [3] process that static document. Handwritten moments are taken as an input to the handwritten recognition system and the input data is the static representation of the handwriting. OCR machines are primarily uses machine printed text and ICR (capital letters)for hand "printed text. The shapes of glyphs and words makes motion capturing easy when

S. C. Satapathy et al. (Eds.): ICETE 2019, LAIS 4, pp. 658–671, 2020.
https://doi.org/10.1007/978-3-030-24318-0_75

taken as input to the technique. The motions captured are the order of drawing the segments, the direction, and the sequence in which the pen is put down and lifted up. With the help of this additional information the accuracy of an end-to-end process can be increased. This technique can also called the “intelligent character recognition”, “on-line character recognition”, “dynamic or real-time character recognition”. On-line handwriting character recognition [4] takes the input from the special digitizer or PDA. The sensor picks up both the pen-tip movements and pen-up/pen-down switching i.e. lifting and putting down of pen. The data collected by the use of this method is called as digital ink. The ink can be considered as a digital representation of handwriting. The signals are converted to the letter codes and can be used in text-processing applications in the computers.

Early versions of character recognition needs to be trained with all images of each character, and has to be operated on one font at a time. But the advanced systems which are used today are capable of achieving a high recognition accuracy for most fonts that are commonly used, and is completed with help of various digital image file format inputs. Some of the systems are also capable of providing outputs of the formatted pages which are approximately same as the original page including images, columns, and other components (non-text).

Humans can easily recognize the handwritten document but the recognition of the same by the computer system becomes difficult for it due to the present of random variations in the noise in image, writing size, fonts and styles.

OCR is a field of research in artificial intelligence, pattern recognition and computer vision. It is broadly used as a method of entering the information from printed paper or a data records such as computerised receipts, invoices, bank statements, business cards, printouts of static-data, mails or any suitable documentation or a passport documents. Digitising printed texts is done by this process for making them electronically edited, stored more compactly, displayed on-line, searched, and can be used in machine processes such as, machine translation, cognitive computing, text-to-speech and data mining.

2 Methodology

Handwritten character or symbol recognition is one of the applications in the pattern classification. Figure 1 shows the block diagram of character recognition.

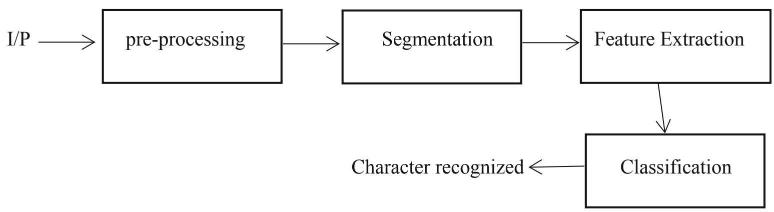

Fig. 1. Block diagram for character recognition

3 Different Handwritten Math Symbol Databases

Handwritten Math Symbol Dataset

This dataset consists of jpg files (45 × 45), dataset does not contain Hebrew alphabet at all. It consists of basic Greek alphabet symbols included are: alpha, beta, gamma, mu, sigma, phi and theta. English alphanumeric symbols are included. All math operators, set operators. Basic pre-defined math functions like: log, lim, cos, sin, [5] tan. Math symbols included are like: \int, \sum, \sqrt, \delta and more (Fig. 2).

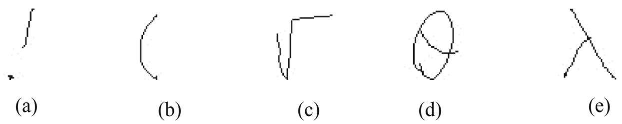

Fig. 2. Sample images from the dataset (a): exclamation mark (b): bracket (c): square-root (d): theta (e): lambda

MNIST

MNIST [6] consists of 60,000 handwritten digits images (64 × 64 × 3) in the training dataset. It also consists of 10,000 images in the test set in gray-scale (28 × 28 × 1). The dataset is easily available for research purpose and is free of cost.

HWRT Database

The dataset consists of 11,081 images. The HWRT database consists of handwritten symbols containing symbols such as all alphanumeric characters, Greek characters, arrows and as well as mathematical symbols like the integral symbol. It is also available easily in jpg format. The size of the jpg images is 156 × 231.

InftyCDB-1

InftyCDB-1 consists of 30 papers in English language which includes of mathematical calculations. It consists of 688,580 alphanumeric character image in 476 pages text document that are recorded along with the character code of the symbol it represents. The links which represent the structure of each word or a mathematical expression are also recorded. InftyCDB-1 [7] can be used as a character database, word database as well as a mathematical expression database. In the InftyCBD-1 database the total number of words images are 108,914 and the total number of images for mathematical expression are 21,056.

HASYv2 - Handwritten Symbol Database

HASY [8] contains 369 symbol classes images (32 × 32). HASY consists of 150,000 instances of handwritten symbols. HASY is a easily available and is free of charge dataset of single symbols similar to MNIST. It contains 168233 instance images of 369 classes.

4 Literature Survey

Kumar et al. [9] proposed "Analytical review of pre-processing techniques for offline handwritten character recognition", it explains Fourier transform method for measuring the shape and pattern to obtain relevant information. The method can easily extract features and classify them. The classification is done using feed forward back propagation neural network. Binarization, character recognition, noise removal, normalization are the preprocessing techniques included along with segmentation. Page segmentation, character segmentation techniques are the segmentation methods used in this paper. Local and global features are extracted from the processed image. The NN classifier is used for classification.

Shi et al. [10] had proposed "Symbol Graph Based Discriminative Training and Rescoring for Improved Math Symbol Recognition". The discriminative training of the exponential weights, the insertion of penalty and graph rescoring are included in the work. In this paper, symbol graph theory of training the exponential weights of each model and inserting penalties is followed. The training is considered for two different areas: maximum mutual information (MMI) and minimum symbol error (MSE). In the post-processing (which is done after training step) trigram-based graph rescoring is performed. The database contains 2500 formulas. The accuracy of 97% for symbol recognition is achieved.

Kasthuri et al. [11] proposed "Noise Reduction and Pre-processing techniques in Handwritten Character Recognition using Neural Networks". The Gabor filtering and noise reduction are the two pre-processing techniques used in this work. The aberrations and non-uniformities are considered as the noise in the process of character recognition. To overcome these issues, it is necessary to perform noise reduction. The process of recognition uses statistical analysis to match between the generated pattern and reference pattern. The dataset used contains 7291 combination of handwritten characters (2 languages) 2549 printed characters. The accuracy achieved is 97%. Jubair et al. [12] proposed "A simplified method for handwritten character recognition from document image". It uses morphological thinning operation as a segmentation technique in this work. The classifier used is KNN. Data-base contains 780 sample images of characters written by different people having different handwritten styles. The accuracy achieved is 95.688%

Hu et al. [13] proposed "HMM-Based Recognition of Online Handwritten Mathematical Symbols Using Segmental K-means Initialization and a Modified Pen-up/down Feature". The work proposed in this paper is recognition work. Hidden Markov Model (HMM) based recognition system is used for recognizing the isolated online handwritten mathematical symbols. Left to right continuous HMM is designed for each symbol class. The symbol recognition includes two steps: symbol segmentation and isolated symbol recognition. K-means produces better initialization using parameters of Gaussian Mixture Models. The pen-up/down gives less accuracy compared to normalized distance to stroke edge features. The database used in this work consists of 22483 sample images. The recognition accuracy obtained for top-1 dataset is 82.9% and top-5 dataset is 97.8%.

Tian et al. [14] proposed “Research on Symbol Recognition for Mathematical Expressions” for recognition of handwritten math symbol in a mathematical expression. Symbol segmentation and symbol recognition are the two phases of the system. The feedback mechanism is proposed for segmentation and recognition of symbols. Symbol segmentation uses the projective features and connected components labelling method for segmentation of an expressions. The directional line element features and peripheral features are extracted from the input symbols. A coarse-to-fine classification strategy is used to recognize symbols with these features. The accuracy achieved is 97.81%. Wang et al. [15] proposed “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”. Normalization is done in pre-processing step. The dataset used is MNIST dataset. Neural net is trained for augmentation and classification. The process is called neural augmentation. Content loss, Style loss via gram matrix and No loss at the end layer are the three augmentation techniques applied in this paper. CNN is used for classification.

Keysers et al. [16] proposed “Multi-Language Online Handwriting Recognition”. Resampling and slope correction are the two pre-processing techniques used in this work. The datasets used are UNIPEN-1, IAM-OnDB and self datsets. The ink is pre-processed. A set of character rules are performed to create a segmentation lattice. Firstly a set of overcomplete potential cut points between the characters are determined using a segmenter. Using the cut points we group set of ink segments into a character hypotheses, this process creates a segmentation lattice. The labelling of the segmentation lattice is accomplished with the help of a classifier by classification of character hypothesis and additional feature functions. The classifier used are HMM and LSTM. The accuracy achieved using UNIPEN-1and IAM-OnDB are 97% and 96%.

Davila et al. [17] had proposed “Using Off-line Features and Synthetic Data for On-line Handwritten Math Symbol Recognition” on-line recognition system to recognize handwritten math symbols that uses off-line features and synthetic data generation. Global features, crossing feature, 2d fuzzy histogram points, fuzzy histogram orientations are the features extracted. Four different classifiers are used AdaBoost C4.5, Random Forest, SVM linear kernel, SVM RBF kernel. Data-base used in this work have two recognition rates top1 and top5. The top-1 and top-5 acquires accuracy of different percentages using the different databases, for AdaBoost C4.5 are 88.4% & 98.7%, for Random Forest are 87.9% 98.4%, for SVM linear kernel are 88.6% 99.1% and for SVM RBF kernel are 89.8% 99.1% respectively.

Rahiman et al. [18] proposed “Recognition of Handwritten Malayalam Characters using Vertical & Horizontal Line Positional Analyzer Algorithm”. Segmentation is done by Line & Character separation. The feature extraction technique used is Horizontal and Vertical Line count and positions. Classifier used is decision Tree for classification. The acquires accuracy is 91%.

Sinha et al. [19] proposed “Zone-Based Feature Extraction Techniques and SVM for Handwritten Gurmukhi Character Recognition”. Zone Centroid Zone method provide better recognition accuracy than Image Centroid Zone. The first attribute according to the writer’s variations in writing style, size, shape, ink colour, ink flow and thickness, digitization imperfections etc. The deficiencies that are fond in the particular method are second attribute for feature extraction. The Gurumukhi dataset of 7000 Gurmukhi character sample images is used in this work. Feature extraction techniques used are Image centroid zone for line segmentation, Zone centroid zone for word segmentation and Hybrid centroid zone for character segmentation. SVM is used for classifier. The accuracy achieved is 95.11%.

Kumawat et al. [20] propose "New Approach of Hand writing Recognition using Curvelet Transform and Invariant Statistical Features". A character recognition system is used for character recognition. The method combines invariant statistical features and curvelet transform. The combined features are used by HMM and SVM to classify the character based on the curvelet transform and invariant statistical features. A 200 samples of two users are taken and the accuracy achieved is 98.92%. Nicolas et al. [21] proposed "Recognition of Handwritten Mathematical Symbols with PHOG Features". The HOG features are generalized to pyramids of HOG features (PHOG). The classifier used is SVM. The CHROME dataset containing 22000 character sample images gets an accuracy of 96%. And 75 handwritten sample images by a different user gets 92% accuracy.

Pradeep et al. [22] proposed "Diagonal Feature Extraction Based Handwritten Character System Using Neural Network". Diagonal features are used for feature extraction and Feed Forward Back propagation Neural Network for classification. Vertical, diagonal and horizontal directions are the different feature attributes used for extracting 54 features from each input character. The different inputs are tested on the Neural network and it performs well. The highest accuracy 98% is obtained by using diagonal orientation for feature extraction this work.

Das et al. [23] propose offline English character recognition model in "HMM based Offline Handwritten Writer Independent English Character Recognition using Global and Local Feature Extraction", to combine the global and local features for classifying the character using hidden Markov model. Data-base contains 13000 samples images of characters written in five different styles for each character collected from 100 writers. The proposed system acquires 98.26% accuracy of 98.26%. Pirlo et al. [24] proposed "Adaptive Membership Functions for Handwritten Character Recognition by Voronoi-Based Image Zoning". The handwritten character recognition system that uses static and dynamic zoning topologies is proposed in this work. The segmentation technique used in this work is Optical image segmentation. This is the technique used for feature extraction and can also be known as Voronoi tessellation.

Malon et al. [25] had proposed "Mathematical symbol recognition with support vector machines" presents a method for the improving classification process. The SVM is used for classification. Multi-class classification by SVM is done by the system utilizing the ranking of alternatives within InftyReader's confusion clusters. Misrecognition rate is reduced by 41% overall. 70,637 samples are taken in the database. Recognition rate of the system is 96.10% obtained without using SVM. The recognition accuracy achieved using SVM by this method is 97.70%. Aradhyal et al. [26] proposed "Robust Unconstrained Handwritten Digit Recognition using Radon Transform". A system is proposed for handwritten digit recognition based on random transform and nearest neighbour algorithm. Two MNIST datasets are used in this work. The first one is of English digit image samples which acquired accuracy of 96.6%, the second one is of Kannada numerals and achieved 91.2%.

Dai Nguyen et al. [27] proposed “Deep Neural Networks for Recognizing Online Handwritten Mathematical Symbols”. The pre-processing technique used is normalization and gradient directional features are extracted. The dataset used is collected from the CHROME which consists of MEs databases containing 120,341 symbol images. Max-out-based CNN is applied to image patterns generated by online patterns. BLSTM is applied to image patterns generated by the original online patterns. Max-out based CNN and BLSTM are also used to combine all the patterns together. Comparing them by MRF and MQDF in a traditional recognition method by experimenting on CROHME database. The classifiers used are CNN and deep max out CNN. The accuracy achieved using BLSTM is 97.61%.

Ratnamala et al. [28] proposed “A novel method for handwritten mathematical document based on equation symbols recognition using K-NN and A-NN classifiers”. The dataset used is self-math document. Filtering and ROI selection is used for pre-processing. Features extracted are statistical and LBP features. The two classifiers KNN and ANN are used for classification. Both the KNN and ANN uses statistical and LBP features for classification of the data into Math or Non-math. The accuracy we achieved by KNN is 96% and ANN is 97%.

Kulkarni [29] proposed “Handwritten Character Recognition Using HOG, COM by OpenCV & Python”. The dataset used the HASY dataset which contains 168233 instances of 369 classes. Gaussian Blurring and Canny Edge Detector are used for data preparation. Segmentation is done through Otsu’s thresholding. HOG features are extracted by de-skewing the images converted to centre of the images. Accuracy of the descriptor is increased power law compression & Square root of the input image. Linear SVM is the classifier used to train the dataset for recognition in the final step. Accuracy achieved by this HCR is 96.56%.

Darmatasia et al. [30] proposed “Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM)”. The dataset used is NIST Special Database 19. The features are extracted by CNN using feature maps. CNN is used as a feature extracted and is constructed using CNNSVM toolbox. A linear SVM combined with CNN is used for classification by L1 loss function and L2 regularization. The accuracy achieved for CNN+SVM is 91.37% while with the original CNN the accuracy achieved is 88.32%.

5 Comparison of Related Works

(See Tables 1 and 2)

Table 1. Comparision between related work for handwritten symbol recognition

S. no.	Author	Title	Preprocessing	Segmentation	Feature extraction	Classification	Accuracy
1.	[9]	Analytical review of preprocessing techniques for offline handwritten character recognition	Binarization, Noise removal, Character Recognition, Normalization	Page segmentation, character segmentation	Local and global features	NN	–
2.	[10]	Symbol graph based discriminative training and rescoring for improved math symbol recognition	Maximum mutual information (MMI) and minimum symbol error (MSE)	MMI and MSE	Weighting	Trigram-based graph rescoring	97%
3.	[11]	Noise Reduction and Pre-processing techniques in Handwritten Character Recognition using Neural Networks	Gabor filtering and noise reduction	Resizing	Fileting	NN	97
4.	[12]	A simplified method for handwritten character recognition from document image	Normalization and binarization	Morphological thinning operation		KNN	95.688
5.	[13]	HMM-based recognition of online handwritten mathematical symbols using segmental k-means initialization and a modified pen-up/down feature	Normalization, edge detection and binarization	Symbol segmentation	Gaussian mixture model, pen up/ pen down features	HMM	83%, 97%
6.	[14]	Research on symbol recognition for mathematical expressions	Normalization binarization and erosion	Feedback mechanism	Projective features	Corse-classification	97.81
7.	[15]	The effectiveness of data augmentation in image classification using deep learning	Normalization threshold thinning	Neural augumentation Content loss, Style loss via gram matrix	Gram matrix, mean, variance	CNN	91.5%
8.	[16]	Multi-Language Online Handwriting Recognition	Resampling, slope correction	Character hypotheses, Segmentation Lattice	Character segmentation	HMMLSTM	97%, 96%
9.	[17]	Using off-line features and synthetic data for on-line handwritten math symbol recognition	Binarization, Normalization and edge detection	Character segmentation	Global features, crossing feature, 2d fuzzy histogram points, fuzzy histogram orientations	AdaBoost C4.5, Random Forest, SVM linear, RBF kernel	98.7% by AdaBoo-top-5 is highest

(*continued*)

Table 1. (*continued*)

S. no.	Author	Title	Preprocessing	Segmentation	Feature extraction	Classification	Accuracy
10.	[18]	Recognition of handwritten Malayalam characters using vertical & horizontal line positional analyzer algorithm	Normalization slant correction noise removal	Line, character separation	Vertical & horizontal line positional analyzer algorithm	Tree classifier	91%
11.	[19]	Zone-Based Feature Extraction Techniques and SVM for Handwritten Gurmukhi Character Recognition	Normalization, binarization and edge detection	Word, Line, character segmentation	Image centroid zone	SVM	95.11%
12.	[20]	New approach of hand writing recognition using curvelet transform and invariant statistical features	Normalization noise removal binarization	Character segmentation	Curvelet transform & invariant statistical features	Combined HMM and SVM	98.92%
13.	[21]	Recognition of mathematical symbol using PHOG features	Binarization thinning edge detection	Symbol character segmentation	PHOG features	SVM	96%
14.	[22]	Diagonal feature extraction based handwritten character system using neural network	Normalization noise removal binarization	Character segmentation	Horizontal, vertical, diagonal features	NN	98%
15.	[23]	HMM based offline handwritten writer independent english character recognition	Binarization Edge detection thining noise removal	Character segmentation	Global and location features using HMM	HMM	98.26%
16.	[24]	Adaptive membership functions for handwritten character recognition by voronoi-based image zoning	Noise removal Normalization	Optical image segmentation	Optimal zoning topology by Voronoi tessellation	Zoning based classifier	–
17.	[25]	Mathematical symbol recognition with support vector machines	Binarization Edge detection	Symbol segmentation	SVM kernel and confusion cluster	SVM kernel	97.07%
18.	[26]	Robust unconstrained handwritten digit recognition using radon transform	Normalization, binarization	Character segmentation	Local. Global features	Random forest	96.6%,91.2%

(*continued*)

Table 1. (*continued*)

S. no.	Author	Title	Preprocessing	Segmentation	Feature extraction	Classification	Accuracy
19.	[27]	Deep neural networks for recognizing online handwritten mathematical symbols	Normalization noise removal	Character segmentation	Gradient directional features	CNN, max-out-CNN, BLSTM	97.61% by BLSTM
20.	[28]	A novel method for handwritten mathematical document based on equation symbols recognition using K-NN and A-NN classifiers	Filtering thinning	ROI	Statistical, LBP features	KNN, ANN	96%, 97%
21.	[29]	Handwritten Character Recognition Using HOG, COM by OpenCV & Python	Gaussian Blurring and Canny Edge Detector	Otsu's thresholding	HOG feature by deskewing	Linear SVM	96.56%
22.	[30]	Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM)	Binarization, Normalization	Character segmentation	Featuring Mapping	CNN + SVM	91.37%, 88.32%

Table 2. Comparision between merits and demerits of the related work

S. no.	Author	Title	Database	Merits	Demerits
1.	[9]	Analytical review of preprocessing techniques for offline handwritten character recognition	Self	Feature extraction step uses the processed image	Need to use all the pre-processing steps to achieve good accuracy
2.	[10]	Symbol graph based discriminative training and rescoring for improved math symbol recognition	Self	Trigram rescoring gets the highest symbol recognition rate	Discriminative training and trigram based graph rescoring is done in post-processing steps
3.	[11]	Noise Reduction and Pre-processing techniques in Handwritten Character Recognition using Neural Networks	2 self dataset	Multiple algorithms is beneficial for character recognition	Language detection is a costlier process and the accuracy decreases when the quality of the input drops
4.	[12]	A simplified method for handwritten character recognition from document image	Self	Less complex, easily implemented and gives high accuracy	Cell value calculation is a complex part

(*continued*)

Table 2. (*continued*)

S. no.	Author	Title	Database	Merits	Demerits
5.	[13]	HMM-based recognition of online handwritten mathematical symbols using segmental k-means initialization	Top-1, top-5	Good accuracy is obtained	Requires lot of time and mental effort
6.	[14]	Research on symbol recognition for mathematical expressions	Self-symbol dataset	Feedback mechanism for segmentation and classification	OCR achieves high accuracy using text but is low for mathematical expressions
7.	[15]	The effectiveness of data augmentation in image classification using deep learning	MNIST	Augmentation techniques for ImageNet dataset gives high accuracy	GAN's and neural augmentation performance is low and takes 3 times more than the traditional augmentation technique
8.	[16]	Multi-Language Online Handwriting Recognition	UNIPEN, IAMOnDB & self	The architecture framework is flexible	Re-use of components across various languages and scripts can be a problem
9.	[17]	Using off-line features and synthetic data for on-line handwritten math symbol recognition	Top-1, top-5	Synthetic data generation for underrated classes can improve average per class accuracy	Ambiguous classes leads to low global accuracy
10.	[18]	Recognition of handwritten Malayalam characters using vertical & horizontal line positional analyzer algorithm	Malayalam characters	Identifies both colored characters and characters with colored background	Due to the similarity in character shapes and character features in Malayalam language the system gives less accuracy
11.	[19]	Zone-Based Feature Extraction Techniques and SVM for Handwritten Gurmukhi Character Recognition	Gumukhi characters	Zone centroid zone and Image centroid zone combination improves the accuracy	Recognition rate depends on SVM parameters
12.	[20]	New approach of hand writing recognition using curvelet transform and invariant statistical features	Self	HMM and SVM kernel are combined to get high efficiency	Feature vectors affects the performance of the system
13.	[21]	Recognition of mathematical symbol using PHOG features	CHROME	PHOG feature extraction techniques along with one against one SVM classifier achieves good accuracy	Complex expressions are difficult to recognize
14.	[22]	Diagonal feature extraction based handwritten character system using neural network	postal address images	Diagonal feature extraction techniques perform better than the conventional horizontal and vertical feature extraction technique	Faces difficulty in using different classifier
15.	[23]	HMM based offline handwritten writer independent english character recognition	Self English character dataset	HMM for some specific characters that have wide range of variation	Low recognition rate for other datasets
16.	[24]	Adaptive membership functions for handwritten character recognition by voronoi-based image zoning	CEDAR and ETL	The segmentation is automatic optimal segmentation and is proven an efficient way	No. of zones should be specified as *a priori*

(*continued*)

Table 2. (*continued*)

S. no.	Author	Title	Database	Merits	Demerits
17.	[25]	Mathematical symbol recognition with support vector machines	InftyReader	SVM with Inftyreader's confusion cluster decreases the misrecognition by 14%	Touching and broken characters treatment is required
18.	[26]	Robust unconstrained handwritten digit recognition using radon transform	MNIST English, kannada numeral	Efficient and robust randon transform method is used in this paper	Range selection is unclear for optimal recognition accuracy
19.	[27]	Deep neural networks for recognizing online handwritten mathematical symbols	CHROME	In offline method MQDF helps to get wide and specific features whereas the BLSTM can access whole document in online method	In the case of mathematical expression the method is difficult to work.
20.	[28]	a novel method for handwritten mathematical document based on equation symbols recognition using K-NN and A-NN classifiers	Self- math document	Initially identifies math symbols and classifies as math and non-math	The overall accuracy achieved is less
21.	[29]	Handwritten Character Recognition Using HOG, COM by OpenCV & Python	HASY	HOG descriptor using edge detection and normalization helps in extracting the features from images of different styles, size and backgrounds	Improper segmentation leads to unambiguous features in feature extraction step
22.	[30]	Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM)	NIST special database 19	Ten folds Cross-validation and document forms containing boundary box and little noise recognition is the work done	Connected character is the problem which is difficult to solve

6 Conclusion

Handwritten document recognition is a complex task to numerous writing styles for distinct person writing styles. The system first identifies the required segment in a handwritten document of characters for segmentation and features are extracted from the segmented character. Characters are recognized from the extracted features. The paper includes the introduction and review on mathematical handwritten character recognition. A literature Survey for pre-processing, segmentation, feature extraction and classification techniques that are effective and efficient for mathematical symbol recognition is briefly explained. The comparison of different papers based on handwritten math symbol recognition is done vividly.

References

1. MacLean S, Labahn G (2015) A Bayesian model for recognizing handwritten mathematical expressions. Pattern Recogn 48(8):2433–2445
2. Singh D, Khan MA, Bansal A, Bansal N (2015) An application of SVM in character recognition with chain code. In: Communication, control and intelligent systems (CCIS). IEEE, pp 167–171

3. Bharambe M (2015) Recognition of offline handwritten mathematical expressions. Structure 34636:97–98
4. Mouchere H, Viard-Gaudin C, Kim DH, Kim JH, Garain, U (2011) Crohme 2011: competition on recognition of online handwritten mathematical expressions. In: 2011 document analysis and recognition (ICDAR). IEEE, pp 1497–1500
5. https://www.kaggle.com/xainano/handwrittenmathsymbols
6. The mnist database of handwritten digits. http://yann.lecun.com/exdb/mnist. Accessed 19 May 2017
7. http://www.inftyproject.org/en/database.html
8. https://zenodo.org/record/259444#.XEStfVwzZPY
9. Kumar G, Bhatia PK, Banger I (2013) Analytical review of preprocessing techniques for offline handwritten character recognition. Int J Adv Eng Sci 3(3):14–22
10. Luo ZX, Shi Y, Soong FK (2008) Symbol graph based discriminative training and rescoring for improved math symbol recognition. In: 2008 IEEE international conference on acoustics, speech and signal processing, ICASSP 2008. IEEE, pp 1953–1956
11. Kasthuri M, Shanthi V (2014) Noise reduction and pre-processing techniques in handwritten character recognition using neural networks. Technia 6(2):940
12. Jubair MI, Banik P (2012) A simplified method for handwritten character recognition from document image. Int J Comput Appl 51(14)
13. Hu L, Zanibbi R (2011) HMM-based recognition of online handwritten mathematical symbols using segmental k-means initialization and a modified pen-up/down feature. In: 2011 international conference on document analysis and recognition (ICDAR). IEEE, pp 457–462
14. Tian XD, Li HY, Li XF, Zhang LP (2006) Research on symbol recognition for mathematical expressions. In: 2006 first international conference on innovative computing, information and control, ICICIC 2006. IEEE, vol 3, pp 357–360
15. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621
16. Keysers D, Deselaers T, Rowley HA, Wang LL, Carbune V (2017) Multi-language online handwriting recognition. IEEE Trans Pattern Anal Mach Intell 39(6):1180–1194
17. Davila K, Ludi S, Zanibbi, R. (2014) Using off-line features and synthetic data for on-line handwritten math symbol recognition. In: 2014 frontiers in handwriting recognition (ICFHR). IEEE, pp 323–328
18. Rahiman MA, Rajasree MS, Masha N, Rema M, Meenakshi R, Kumar GM (2011) Recognition of handwritten Malayalam characters using vertical & horizontal line positional analyzer algorithm. In: 2011 3rd international conference on electronics computer technology (ICECT). IEEE, vol 2, pp 268–274
19. Sinha G, Rani A, Dhir R, Rani MR (2012) Zone-based feature extraction techniques and SVM for handwritten gurmukhi character recognition. Int J Adv Res Comput Sci Softw Eng 2(6)
20. Kumawat P, Khatri A, Nagaria B (2013) New approach of hand writing recognition using curvelet transform and invariant statistical features. Int J Comput Appl 61(18)
21. Jimenez ND, Nguyen L. Recognition of Handwritten Mathematical Symbols with PHOG
22. Pradeep J, Srinivasan E, Himavathi S (2010) Diagonal feature extraction based handwritten character system using neural network. Int J Comput Appl 8(9):17–22
23. Das RL, Prasad BK, Sanyal G (2012) HMM based offline handwritten writer independent english character recognition using global and local feature extraction. Int J Comput Appl 46 (10):45–50

24. Pirlo G, Impedovo D (2012) Adaptive membership functions for handwritten character recognition by voronoi-based image zoning. IEEE Trans Image Process 21(9):3827–3837
25. Malon C, Uchida S, Suzuki M (2008) Mathematical symbol recognition with support vector machines. Pattern Recogn Lett 29(9):1326–1332
26. Aradhya VM, Kumar GH, Noushath S (2007) Robust unconstrained handwritten digit recognition using radon transform. In: 2007 signal processing, communications and networking, ICSCN 2007. IEEE, pp 626–629
27. Dai Nguyen H, Le AD, Nakagawa M (2015) Deep neural networks for recognizing online handwritten mathematical symbols. In: 2015 3rd IAPR Asian conference on pattern recognition (ACPR). IEEE, pp 121–125
28. Ratnamala SP (2016) Shilpa proposed a novel method for handwritten mathematical document based on equation symbols recognition using K-NN and A-NN classifiers. International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) 23(6). (SPECIAL ISSUE), ISSN 0976-1353
29. Kulkarni RL (2017) Handwritten character recognition using HOG, COM by OpenCV & Python. Int J 5(4)
30. Fanany MI (2017) Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM). In: 2017 information and communication technology (ICoIC 2017). IEEE, pp 1–6