



Machine learning models for mathematical symbol recognition: A stem to stern literature analysis

Vinay Kukreja¹ · Sakshi¹

Received: 14 June 2021 / Revised: 31 October 2021 / Accepted: 9 February 2022 /
Published online: 31 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Given the ubiquity of handwriting and mathematical content in human transactions, machine recognition of handwritten mathematical text and symbols has become a domain of great practical scope and significance. Recognition of mathematical expression (ME) has remained a challenging and emerging research domain, with mathematical symbol recognition (MSR) as a requisite step in the entire recognition process. Many variations in writing styles and existing dissimilarities among the wide range of symbols and recurring characters make the recognition tasks strenuous even for Optical Character Recognition. The past decade has witnessed the emergence of recognition techniques and the peaking interest of several researchers in this evolving domain. In light of the current research status associated with recognizing handwritten math symbols, a systematic review of the literature seems timely. This article seeks to provide a complete systematic analysis of recognition techniques, models, datasets, sub-stages, accuracy metrics, and accuracy details in an extracted form as described in the literature. A systematic literature review conducted in this study includes pragmatic studies until the year 2021, and the analysis reveals Support Vector Machine (SVM) to be the most dominating recognition technique and symbol recognition rate to be most frequently deployed accuracy measure and other interesting results in terms of segmentation, feature extraction and datasets involved are vividly represented. The statistics of mathematical symbols-related papers are shown, and open problems are identified for more advanced research. Our study focused on the key points of earlier research, present work, and the future direction of MSR.

Keywords Convolutional neural network · Handwritten mathematical symbol recognition · Machine learning · Mathematical symbol recognition · Optical character recognition · Support vector machine · Segmentation

✉ Sakshi
asakshi541@gmail.com

Vinay Kukreja
onlyvinaykukreja@gmail.com

¹ Chitkara University Institute of Engineering and Technology, Chitkara University, PunjabRajpura, India

1 Introduction

Although an enormous amount of work has been done in mathematical symbol recognition, a more precise recognition of mathematical symbols is still a big concern. The applications of handwritten input in electronic devices are massive; that is why handwritten symbols or character recognition technology have captured the interest of academicians and industry experts [19]. Optical Character Recognition handles the tasks related to handwritten symbols or character recognition efficiently as it has high skills to do so but with a strong dependency on contextual information, which is arduous to apply on handwritten mathematical formulas.

According to the research work of the authors [14], the increased accessibility of pen and touch-based devices such as smartphones, tablets, personal computers, and smartboards has proliferated the interest in mathematical symbol recognition. Thus, it has been growing. Still, existing systems are away from perfection because of complications arising from the typeset with many similar-looking symbols or characters and the two-dimensional nature of mathematical input. Recently, handwriting recognition used encoder-decoder models for image to text transformation [73, 82]. This recognition trend has evolved a lot since the time this field has passed been embarked for exploration.

Handwriting Recognition is divided into two categories: offline and online. “Off-line” encompasses the images of the written text papers taken by optical image scanning or Intelligent Word Recognition, whereas “On-line” relate to the movement of the pen tip, which is in contact with the pen-based computer screen surface [45].

Mathematics is free from racial, national, and cultural constraints as the “unique international scientific language.” For all kinds of documents, mathematical formulas are an essential part. The development of information requires that types of document image processing systems be mature and perfect, which can handle characters, tables, and images well [61]. After this, the challenging domain captivated more attention and concentration, and in the ‘90s, the hot topic of pattern recognition is mathematical formula recognition. Author [8] conducted their study to survey different techniques for symbol segmentation and recognition for online and offline math recognition.

A mathematical symbol recognition system must be powerful enough to manage various specific symbols.

The current literature highlights that Machine Learning (ML)-based models are used for recognition and have been extensively used. The recent studies witness the ML-based recognition models to be achieving significant accuracy rates. For instance, research by [52] uses a K-NN classifier and finally achieved 90% accuracy. In contrast, another study [46] uses the CNN technique as a recognition technique and achieves a competitive accuracy of 93.5%.

All the above considerations and significant stances of research have inspired us to summarize and extract the findings of the current research on the ML approach for mathematical symbol recognition. To this goal, this review describes the SLR of the studies published until 2021 year, following the SLR guidelines. Furthermore, it has been believed that an appreciable amount of work has been done in this domain in the past few years. In this SLR, the authors cover seven research questions. Studies to conduct this SLR are extracted from the most populated digital libraries, i.e., Scopus, Springer, IEEE Xplore, and Google Scholar, based on inclusion/exclusion factors and quality assessment scores.

The order of the paper is described in section 2; the authors detail the entire review process, discussing the procedural steps involved in the conduction of this SLR. In section 3, the authors focus on the necessary preliminaries for this review. In section 4, a detailed and particularized all necessary observations to answer the formulated research questions. Each

discussion about the research questions is the result of extensive analysis and data extraction. In section 5, the authors have tried to summarize the gist of observations under the heading “summary of findings”. In section 6, the conclusions and recommendations for future work have been offered. The entire organization of the paper is presented in Fig. 1.

1.1 Motivation for research

1. Recognizing handwriting has been an evolving research domain. The field of mathematical symbols and expression recognition is the prominent research domain that has been a proactive research theme for the past 40 years. This systematic review objectifies to brief all the recognition models that could channelize and explore the literature along crucial dimensions and directions.
2. The landmark survey in this domain was published in 2012 by the authors [78]. Therefore, a gap of 9 years has witnessed the emergence and shift to other implementation models and recognition techniques. We recognized the requirement of a comprehensive literature survey after considering progressive research to recognize handwritten math text. Therefore, we realized the need to compile the existing literature based on a complete and systematic search to present the research challenges for future study.
3. In the context of scientific motivation, the goal of recognizing handwritten MEs is representative of pattern recognition in two dimensions. Therefore, the recognition technique for MEs can be used for others like flow charts, diagrams, etc. For academic motivation, recognizing handwritten math notations can help both the teachers and the students for information retrieval for identifying mathematical notation. A set of relevant documents can be obtained using ME as a query.
4. Another main motivation for reviewing the recognition ability and systems that identify mathematical text is that MEs and other related text can be easily entered into the machines and devices. Inputting directly from a handwritten source is far more realistic than using pre-existing formats for math descriptions like TEX, MathML. Once such an expression gets correctly interpreted, it can be pasted to an article, evaluated by a mathematical program, or

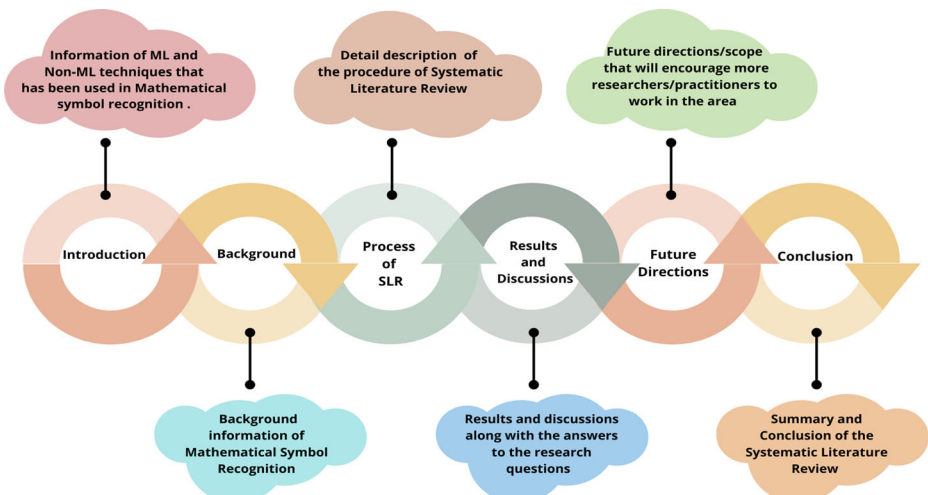


Fig. 1 Organization of the paper

plotted as a graph. Thus, recollecting and compiling the literature based on recognizing online handwritten mathematical symbols is the chosen subject of this paper.

1.2 Our contributions

- A systematic investigation has been conducted to study various recognition models, classifying the work based on the kind of technique used for recognition (ML/Non- ML).
- The identified procedures and techniques have been compared and categorized based on their kind, segmentation, and feature extraction techniques. The datasets deployed to implement the recognition models have been identified, and corresponding metrics used for accuracy have also been extracted. Further, the inclusion relationship among different metrics and the size of the dataset have been shown in Section 4.6.
- Statistical results (yearly publications, type of publications, dataset frequency analysis) related to handwritten mathematical symbol recognition papers have been presented.
- Future research directions relating to the current compiled literature on handwritten mathematical symbol recognition have been widely depicted.

1.3 Related surveys

Earlier surveys by Chan [8] and Zanibbi [78] have been very innovative. Still, as the research has consistently grown in mathematical symbol recognition, this required a systematic literature survey, which narrows down this field's existing research. Chan has reviewed the recognition processes like segmentation and feature extraction details based on the online and offline handwritten text classification. Zanibbi has explored and studied the research works based on representation models and identified the challenges for recognizing and retrieving MEs. It identified the critical problems in math recognition: detecting expressions, detecting and classifying symbols, analyzing symbol layout, and constructing a representation of meaning. Another survey in the list is added by Firdaus [22], which focuses on the merits and demerits of several recognition techniques for mathematical symbols. The recent survey report by Sakshi [62] undermines the statistical extractions derived from the combined literature of mathematical symbols and expressions, leaving a decent scope for attention and specification for literature investigation purely based on isolated symbol recognition.

This research review will be a complete value addition to the existing literature on the handwritten mathematical symbol, as it's the first-ever systematic survey that aims to compile the literature from stem to stern concerning the recognition models trends in the context of symbol recognition and observes the research challenges concerning root problems.

2 Background

2.1 Mathematical symbol recognition

ME recognition mainly includes two steps: symbol recognition and structural analysis. The spatial relationships between symbols determine the structural analysis and interpret the expression [1, 4], (Carlos Ramirez-Pina et al., 2019). The basis of structural analysis is symbol

recognition. There are two phases involved in symbol recognition, which are named symbol segmentation and isolated symbol recognition. The online ME is a set of strokes of input data, and the inclusion of more than one stroke forms a mathematical symbol. Converting stroke series into a set of symbols is called symbol segmentation, further classified in the isolated symbol recognition stage [8, 29].

The symbols that come under MEs are mathematical symbols, multi-front characters, and Greek letters. To extract the symbol from a ME correctly is not an easy task. The process of extracting mathematical symbols from the ME is termed mathematical symbol recognition, requiring specific necessary steps. There are three main stages in most recognition systems: Preprocessing, feature extraction, and matching. The processing environment for subsequent features provides in the preprocessing stage. The most crucial step in pattern recognition is feature extraction. The static properties and dynamic are factors for feature analysis like descender or no descender, dot, or no dot (Tappert, 1990). The selection of the most likely candidate for an unknown input character comprises the process called matching. [42].

2.2 Types of mathematical objects

Mathematical objects are used for communication used by a mathematician in mathematical language. The kernel of a mathematical document is a mathematical object. It can be a set of elementary mathematical objects arranged according to mathematical language grammar rules, as shown in Fig. 2. A small entity with a sense and is indivisible in other mathematical objects is an elementary mathematical object. Mathematical objects or sets of mathematical objects, possibly interconnected or arranged by operators, are MEs [31].

2.3 Types of mathematical symbol recognition

There are generally two categories: online and offline recognition methods, Handwritten character pattern recognition, or mathematical symbol recognition [43]. Earlier online patterns

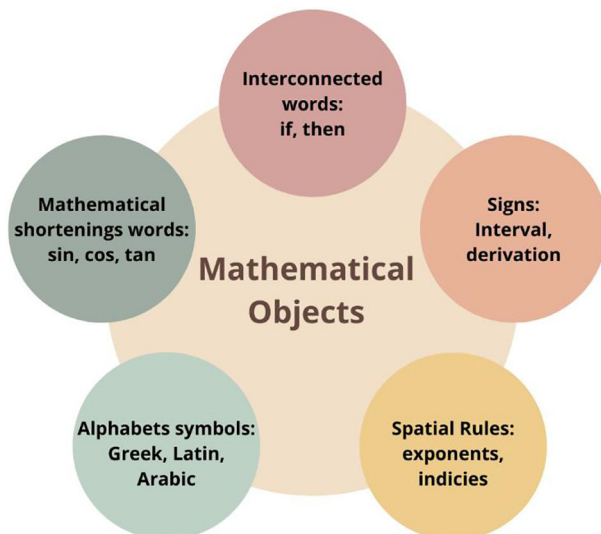


Fig. 2 Types of mathematical objects

are recognized using the Markov Random Field as these online patterns are the strokes' time order. Pen-based or touch-based devices are used to record the coordinates of pen-tip or fingertip trajectory, which is known as stroke [83], and to capture the 2D images of offline patterns which are scanned using the scanner or camera are further recognized by Modified Quadratic Discriminant Function and directional feature extraction. If we can discard the temporal and structural information from the pattern, then the method used to recognize offline patterns can identify online patterns.

The forward-backward neural network-based algorithms like Bidirectional Long Short Term Memory (BLSTM) networks with a Connectionist Temporal Classification output layer have exhibited outperform state-of-the-art Hidden Markov Model-based systems to recognize online handwriting. Combining the online and offline information to extract the set of hybrid features and using the Hidden Markov Model and BLSTM networks for classification is recently proposed by authors [3]. Raw images as local offline features employed by BLSTM alongside the pen-tip curve extensively outperformed Hidden Markov Model in the symbol recognition rate.

2.4 Characteristics of mathematical symbols

According to the relationships defined among the symbols used in the mathematical expressions, there are several characteristics of symbols displayed as follows:

- The scope of fractions, roots, and matrices is determined by its “core symbol”, which gives the meaning of sub-expressions such as “-”, “ $\sqrt{\quad}$ ”, “(“; [“; { “; P”, etc.. These core symbols can be further into two types: fence symbols and binding symbols [9].
- There are fence symbols, which serve as fences around expression, to be treated as a unit. For instance, the enclosed units are grouped into a unit with the help of parentheses. For example, “(ab + c)” is a unit that holds the sum of the product of a and b with c.
- Some symbols are termed as binding symbols that purposely bind together several entities of expression and give meaning to expression as a whole. Like the fraction line $\frac{x}{y}$, \sqrt{x} and Σx_i , dominate their neighboring expressions but bounds several entities and represents it as a single unit. For example $\sum_{i=1}^n X_i$, this represents the sum of n terms, i.e., $X_1 + X_2 + X_3 + \dots + X_n$.
- According to [54], the scope of scripts and limits is determined by the previous and subsequent symbols on the same line. For example, pre-super script, upper script, superscript, pre-subscript, lower script, and subscript are the possible positions/ scripts around the symbol.
- Further, there are operator symbols, which include an operator for evaluation like +, -, /, ÷, *. These operator symbols dominate and operate their operands. As mentioned in the findings of authors [8], the operators can be classified into types, namely, explicit and implicit operators. Explicit operators are operator symbols, and implicit operators include spatial operators, which provide for geometrical properties defined in physical space.

3 Systematic review process

The procedure of the SLR follows the strategy recommended by Kitchenham [39]. It contains the arrangement, the execution, and the result analysis description. The first stage in the

procedure of SLR is the planning stage, in which we classify and build up the review protocol that comprises the steps presented in Fig. 3.

After the planning stage, a series of measures are carried out wherein the first step, the research questions are posed associated with the SLR's objectives. In the second stage, the reviewers designed search tactics to determine those potential studies to answer the formulated research questions. The primary task of this stage is to recognize and corroborate the search terms (formation of search string) and the literature resources for extracting our set of studies for this review. According to our framed research questions, phase 3 of this SLR deals with extracting the relevant and appropriate studies from chosen digital data sources. The fourth step includes the inclusion/exclusion parameters to conclude the studies included and excluded from SLR.

Further quality assessment criteria are taken to determine individual studies' strength and quality by forming the quality assessment questionnaire and using the fuzzy linguistic variable's value to access the studies. In the second last step, research questions formulated in this study are answered based on the information gathered and stored in the data extraction form. The final step comprises analyzing the extracted data collected stored in the data extraction form in accordance with our designed research questions described in Table 1. The flow sequence of the SLR procedure is depicted in Fig. 3. The review protocol development is an essential step in an SLR to minimize the risk of research bias in the SLR.

3.1 Research questions

This systematic analysis evaluates the pragmatic evidence from different studies using ML and non-ML techniques for mathematical symbol recognition. The formulated research questions have been presented in Table 1.

To cover all the aspects that have to be investigated in terms of mathematical symbol recognition, the review group has chosen the above research questions. The authors indeed focused on ML techniques, corpus, accuracy measures employed, and publication sources for mathematical symbol recognition.



Fig. 3 Systematic Literature Review (SLR) Procedure

Table 1 Research Questions

Id	Research Questions	Motivation
RQ1	Which ML/non-ML techniques are employed in mathematical symbol recognition?	To inspect what are the variant of ML and Non-ML techniques have been used in mathematical symbol recognition and to showcase percentage distribution of ML and Non-ML techniques
RQ2	Which are corpus mainly used for experimentation in mathematical symbol recognition?	To analyze different datasets/corpus has been employed in mathematical symbol recognition selected studies.
RQ3	What are the majorly used accuracy metrics for mathematical symbol recognition?	To explore the enormously employed accuracy parameters for mathematical symbol recognition
RQ4	What are segmentation/feature extraction techniques prominent for mathematical symbol recognition?	To investigate segmentation/feature extraction techniques that the chosen studies have majorly used for mathematical symbol recognition.
RQ5	What are trend analysis statistics pertinent to the performance of ML techniques with other ML and non-ML techniques?	For comparing the performance of ML and non-ML techniques based on the accuracy parameter and accuracy values achieved.
RQ6	Is the size of the dataset a factor influencing the accuracy rate of mathematical symbol recognition?	To examine whether the size of the dataset is a dependable parameter for affecting the accuracy rate of mathematical symbol recognition
RQ7	What are the predominant publication channels for research communications in mathematical symbol recognition?	This research question aims to probe the cardinal publications channels for research communications in mathematical symbol recognition

3.2 Search tactics

Our foremost searching approach is to split into two stages for identifying and downloading studies a) Primary search b) Secondary search.

Following are the procedural steps for primary search:

- A. Recognize the keywords from the research questions
- B. Recognize synonyms and related words in step A
- C. Use Boolean operations such as OR for synonyms and related words or terms and Boolean AND to join significant terms.

The review group investigated the bibliographical content of the chosen primary set of studies to look for those studies that could have been missed or relatively ignored in the primary search. This falls under our secondary search phase. The string used for search is:

“mathematical symbol recognition” OR “Handwritten Formula Symbol Recognition” OR “Math symbol recognition” OR “Math symbol retrieval” OR “Mathematical symbol retrieval” OR “handwritten mathematical symbols.”

Using related terms and synonyms of associated terms using the Boolean operator OR and uniting the main terms via the Boolean operator AND complicated search terms have been included in similar systematic reviews. The search string is thus framed for fetching all accessible and relevant studies in the databases. The articulated search string based on the Boolean operators OR and AND together with all possible synonyms is run on the selected digital libraries. For example, we have combined the synonyms of recognition and retrieval using the OR operator because various studies use mathematical symbol recognition, while

others use MS retrieval. In the same way, searches have been done for “maths” and “mathematical” terms. The authors have used maths* and mathematical*, which means if the words used in the individual studies are maths, mathematical, and mathematics, they are preferred and chosen by the framed search string.

The following are the sources of searching and selecting the primary studies for this literature review:

- IEEE Xplore
- Scopus
- Springer
- Google Scholar

The above digital databases are chosen for data collection because, according to the AMSTAR checklist, at least two databases, as well as at least one supplemental search, should be searched (Shea et al., 2017). Furthermore, the most recent PRISMA 2020 [55] standards suggest that research for SLR should be retrieved from multiple databases. To mitigate the consequences of varying coverage between particular databases, specific authors in other domains generally recommend a minimum of two databases [26, 51]. As a result, the authors could complete their search on IEEE, Scopus, and Springer, which covered the majority of the relevant studies. To avoid missing any significant papers for the review, the authors expanded their Google Scholar search to include all relevant and semi-related studies. Furthermore, these four digital libraries drew a sizable number of relevant studies.

The studies until 2021 have been gathered. The individual digital library has fetched a specific number of articles according to the search string. The total sum of 648 studies has been filtered based on search terms. The titles and abstracts of these 648 studies have been investigated, resulting in 185 studies. Thereafter the duplicate studies have been removed and finalized 94 studies are chosen after accomplishing the primary search phase. The secondary search for more relevant studies is performed by considering the references of the selected set of studies. This secondary search phase led to 12 significant additional studies (missed in the initial search). Therefore, the total comprehensive studies chosen for this SLR are 106 based on our primary and secondary search phrases. Figure 4 depicts the steps used for determining and selecting the studies for this review.

3.3 Study selection

The selection of studies is initially based on the relevance of their titles and abstracts. The review group then decided to employ two more phases to get more propitious and consistent literature. The two phases are namely: inclusion-exclusion and quality assessment. Firstly, the inclusion and exclusion criteria are applied to the collected corpus of studies, which helped narrow our decision process while selecting the studies for this review. Secondly, quality assessment criteria are implemented for further filtration of chosen studies. The following (Table 2) are the factors defining our inclusion/exclusion criteria that have been deliberately deployed.

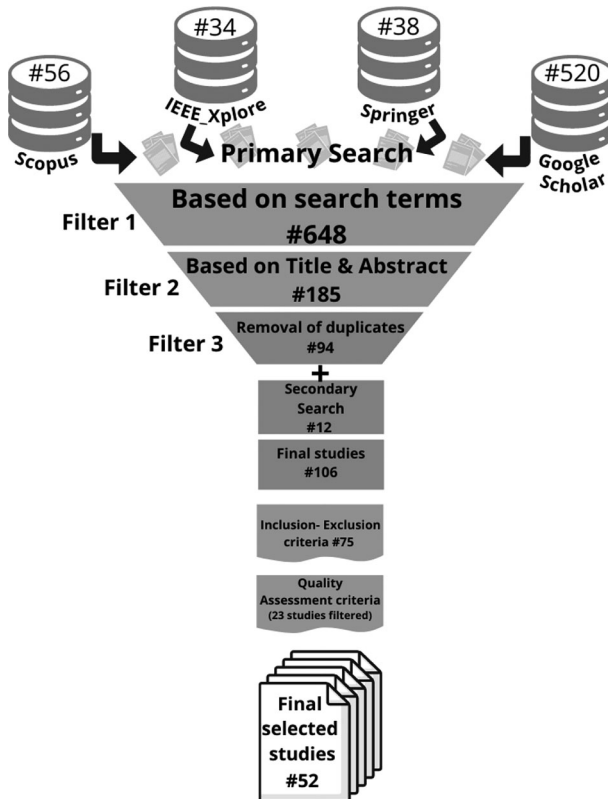


Fig. 4 Stepwise selection and filtering of studies

3.4 Quality assessment criteria

To determine the integrity and relevance of the studies considered, quality assessment is evaluated. An assessment criterion is another method to select the studies from the raw studies corpus. Firstly, the inclusion and exclusion criterion is applied to the studies, and further questions are framed to assess the quality of the studies to achieve potential studies for

Table 2 Inclusion-Exclusion Factors

Inclusion	Exclusion
<ul style="list-style-type: none"> • Studies that are only focusing on Mathematical symbols. • Studies that used ML-based recognition models for mathematical symbol recognition. • Studies that used the combination of both ML and non-ML based recognition models for mathematical symbol recognition • For studies that are published both in journals and conferences • Studies that used accuracy metrics to justify or to compare their research or survey. 	<ul style="list-style-type: none"> • Studies that do not focus on mathematical symbol recognition instead focus on symbol recognition • Studies that focus on the recognition of MEs rather than mathematical symbols. • Review studies (i.e., without experimental investigations)

literature review. Studies found with low value are excluded from the studies. The question considered to conduct the quality assessment is given below:

- Q1. The research aim is unambiguously affirmed or not?
- Q2. Are the estimation methods well defined and deliberated?
- Q3. Is the experiment applied to sufficient datasets?
- Q4. Is the accuracy of estimations measured?

Related work of quality assessment criteria is studied, and various meetings and discussions are conducted between the authors. After the successful meetings, we finalized the fuzzy linguistic variable method to assess the quality of the selected studies. Fuzzy linguistic variables are taken in this study to score considered quality assessment questions, and this SLR used the score from 0 to 1 for each question. This concludes that finalized potential studies range from 0 to 1 as a score value for each study. Score 0 is not considered an efficient way to measure a specific quality question. In our view, it's not fair to weigh them in this order, and this is the idea behind the fuzzy linguistic variables. For example, we have a quality question: "Is the experiment applied on a sufficient number of datasets?" Let us consider that any study has five datasets, and we assigned a precise value of 1 to this question, but questions arise what value is assigned to question if we have 2 or 3 datasets. Will we assign value 0? To avoid giving a crisp value to the question, fuzzy linguistic variables are considered a novel idea. To solve this problem, exact values are avoided in this study, and the fuzzy linguistic variables approach is used. The score range of each question is mentioned in Table 3.

Labeling and the total score for each considered question are shown in Table 4.

In this study, five linguistic variable values are considered, which are mentioned in Table 4. Studies having value no and rarely are excluded from the corpus as these values are regarded as low score values. We have further discarded 23 studies as a result of a quality assessment score. Studies having score value partly, mostly, and yes are considered confirmed studies to conduct this SLR. Studies that are scored as "yes" are judged as high-quality studies, and studies with a score value of partly and mostly are considered partial good studies. The score of each study is shown in Table 5.

The studies have a high impact that can be highlighted to our readers. For example, we can consider high-impact studies as those that answered most of the research questions. After all the steps followed in selecting the final corpus, we finalized 52 relevant studies to conduct this SLR. At the end of the paper, Table 13 in Appendix A depicts the complete form of abbreviations.

3.5 Data extraction

For every research question formulated for this review, separate files of excels are prepared. For instance, the count of publications per year had individual records. In contrast, different data extraction files manage the technique used, the dataset deployed, and other associated factors. One of the authors extracted the raw data from the selected studies. Simultaneously,

Table 3 Range and Linguistic score of Quality Assessment

Range	0	0.1–0.3	0.4–0.6	0.7–0.9	1.0
Score	No	Rarely	Partly	Mostly	Yes

Table 4 Cumulative Assessment Score with Labels

0(No)	0.1–1.1 (Rarely)	1.2–2.2 (Partly)	2.3–4.1 (Mostly)	4.3 or more (Yes).
-------	------------------	------------------	------------------	--------------------

the other is arranged in order as per the requirement of this review. The third author thoroughly revised and validated the authenticity of the data arranged after extraction.

Final studies of this SLR with the research question they addressed are shown in Table 6. Each study doesn't need to provide answers to every research question formulated in SLR. The chosen final list of studies is selected and scored based on two factors: a quality assessment score and the number of questions addressed by the study. Studies are scored from 0 to 10, 0 is taken as a minimum score, and 10 is taken as the maximum score. According to score value studies, credibility is evaluated and represented in Table 5. In this SLR, seven research questions are formulated, and on these questions, one study can get a maximum of seven scores, i.e., one point for each research question. In the quality assessment score, one study can reach a maximum score of three as if the study value is “yes” it has been scored as 3, for “mostly” it is scored as 2 and “partly”, it is scored as 1. Considered studies are scored out of 10 based on the mentioned parameters. This idea of assigning a score to each study is to recommend quality studies to our readers.

3.6 Data synthesis

The data synthesis step aims to identify and group all the confirmation from the chosen set of potential studies related to our research questions. A small piece of information might be considered little evidence, but this information collection can strengthen a point. To synthesize our result data, most tables are used, and to use the results, vote-counting measures are taken for comparative analysis.

Table 5 Quality assessment scores and their linguistic variable

Studies	Quality Assessment Score	Studies	Quality Assessment Score	Studies	Quality Assessment Score
S1	Mostly	S18	Mostly	S36	Mostly
S2	Mostly	S19	Mostly	S37	Mostly
S3	Mostly	S20	Mostly	S38	Partly
S4	Mostly	S21	Yes	S39	Yes
S5	Mostly	S22	Partly	S40	Partly
S6	Mostly	S23	Yes	S41	Mostly
S7	Mostly	S24	Mostly	S42	Yes
S8	Mostly	S25	Yes	S43	Mostly
S9	Mostly	S26	Yes	S44	Mostly
S10	Mostly	S27	Mostly	S45	Mostly
S11	Yes	S28	Mostly	S46	Mostly
S12	Mostly	S29	Partly	S47	Mostly
S13	Mostly	S30	Mostly	S48	Mostly
S14	Mostly	S31	Mostly	S49	Mostly
S15	Partly	S32	Yes	S50	Mostly
S16	Mostly	S33	Mostly	S51	Mostly
S17	Yes	S34	Mostly	S52	Partly

Table 6 SLR's research questions addressed by the selected studies

STUDIES	References	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6	RQ7
S1	[15]	Yes	Yes	No	Yes	Yes	Yes	Yes
S2	[38]	Yes	Yes	No	Yes	Yes	Yes	Yes
S3	[11]	No	Yes	No	Yes	Yes	Yes	Yes
S4	[44]	Yes	No	No	Yes	Yes	Yes	Yes
S5	[71]	Yes	Yes	No	Yes	Yes	Yes	Yes
S6	[56]	Yes	Yes	Yes	No	Yes	Yes	Yes
S7	[77]	Yes	No	No	Yes	Yes	Yes	Yes
S8	[4]	Yes	Yes	No	Yes	Yes	Yes	Yes
S9	[58]	Yes	Yes	No	No	Yes	No	Yes
S10	[17]	Yes	Yes	No	Yes	Yes	No	Yes
S11	[27]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S12	[60]	Yes	No	Yes	Yes	Yes	Yes	Yes
S13	[6]	Yes	No	No	Yes	Yes	Yes	Yes
S14	[7]	Yes	No	Yes	Yes	Yes	No	Yes
S15	[81]	No	No	No	Yes	Yes	Yes	Yes
S16	[52]	Yes	Yes	No	Yes	Yes	Yes	Yes
S17	[1]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S18	[59]	Yes	Yes	No	Yes	Yes	Yes	Yes
S19	[47]	Yes	Yes	No	Yes	Yes	Yes	Yes
S20	[21]	Yes	Yes	No	Yes	Yes	Yes	Yes
S21	[14]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S22	[32]	No	No	No	Yes	Yes	No	Yes
S23	[53]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S24	[23]	Yes	No	No	Yes	Yes	Yes	Yes
S25	[2]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S26	[20]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S27	[69]	Yes	No	No	Yes	Yes	Yes	Yes
S28	[34]	Yes	Yes	No	Yes	Yes	Yes	Yes
S29	[37]	Yes	No	No	No	Yes	Yes	Yes
S30	[3]	No	Yes	Yes	Yes	Yes	Yes	Yes
S31	[41]	Yes	Yes	No	Yes	Yes	Yes	Yes
S32	[29]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S33	[49]	Yes	Yes	No	Yes	Yes	Yes	Yes
S34	[50]	Yes	Yes	No	No	Yes	Yes	Yes
S35	[25]	Yes	Yes	No	Yes	Yes	Yes	Yes
S36	[63]	Yes	No	No	Yes	Yes	Yes	Yes
S37	[64]	Yes	No	No	Yes	Yes	Yes	Yes
S38	[76]	Yes	No	Yes	No	Yes	No	Yes
S39	[61]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S40	[24]	Yes	No	Yes	No	Yes	No	Yes
S41	[48]	Yes	Yes	No	Yes	Yes	Yes	Yes
S42	[19]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S43	[68]	Yes	Yes	No	Yes	Yes	Yes	Yes
S44	[28]	Yes	No	No	Yes	Yes	Yes	Yes
S45	[16]	Yes	Yes	Yes	No	Yes	Yes	Yes
S46	[67]	Yes	No	No	Yes	Yes	Yes	Yes
S47	[75]	Yes	No	No	Yes	Yes	Yes	Yes
S48	[74]	Yes	No	No	Yes	Yes	Yes	Yes
S49	[35]	Yes	No	No	Yes	Yes	Yes	Yes
S50	[72]	No	Yes	No	Yes	Yes	Yes	Yes
S51	[40]	Yes	No	No	Yes	Yes	Yes	Yes
S52	[33]	Yes	Yes	No	No	Yes	Yes	Yes

3.7 Threat to validity

3.7.1 Study selection and discrimination

In this SLR, we selected 52 studies, and to finalize these studies search string is tried on different platforms according to our formulated research questions. There may be a chance some reviews do not include the string keywords in their Metadata, which creates the possibility of rejecting that study. To choose the studies, we also refer to the references used by different authors to include essential studies in the corpus. However, still, there is the possibility that some relevant studies are missed, which can be considered a threat.

3.7.2 Subjective quality assessment

The specified criteria settled by quality assessment led to the discarding of studies that don't fulfill the set criteria bounds. Thus there is a decent probability that prime studies in this field might not have been considered, owing to their un-fulfillment to the settled criteria by the quality assessment phase. This could be a potential threat. Using a fuzzy linguistic variable gauges each study's score value and excludes the studies with the low score value. Scoring of the selected studies is done by the three authors involved in this SLR. This could also be a threat. Credibility score related to each study is shown in Fig. 5, and it depicts that studies S11, S17, S21, S23, S25, S26, S32, S39, S42, S43, have a high credit score, whereas studies like S15, S22, S29, S38, S40, S52 having a low score.

4 Results and discussions

In this section, results are obtained from our final selected research questions.

4.1 RQ1 which machine learning /non-machine learning techniques are employed in mathematical symbol recognition?

This SLR analyzes the ML and non-ML techniques used in all the selected studies, as shown in Fig. 6. On examining the trend of strategies implemented for the recognition tasks, it is observed that the ML techniques are comparatively more frequently deployed than the non-ML methods.

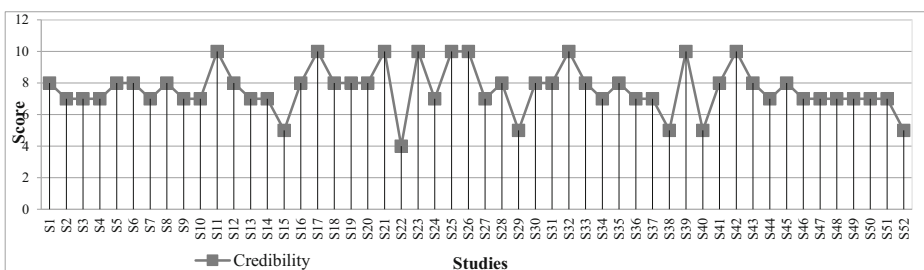
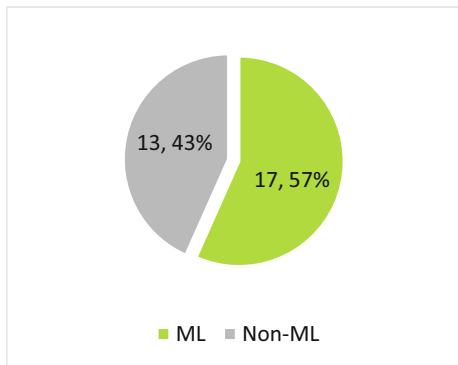


Fig. 5 Overall credibility of the selected studies

Fig. 6 ML and non-ML techniques analyzed by SLR



The further analysis of ML techniques revealed that SVM is a widely used ML model implemented in almost 21% (#11) of the total studies. The success of SVM can be attributed to the fact that it performs effectively with distinct separation margins. Furthermore, mathematical symbols are carved in two dimensions, and SVM effectively deals with a wide range of dimensions and samples. It is also memory efficient from an algorithmic standpoint. The use of SVM, according to the authors, has resulted in competent recognition rates (presented later in the study in Table 11). As a result, SVM has become increasingly popular for symbol recognition tasks.

The following most applied ML methods are K-nearest neighbor and Convolutional neural network (CNN). The ML methods have been deployed in about five studies individually, constituting about 9% (each) of the chosen studies. The other neural network models like Artificial neural networks and Recurrent neural networks have also been employed in about a few studies, ranging from approximately 2 to 3% of selected studies. BLSTM, part of Feedforward neural networks, has been used in 5% (#3) of studies.

There have also been observations regarding deep learning and CNN architectures deployed in the selected studies. The authors have extended efforts to present a timeline of varied deep learning architectures in different studies chosen over the years (Fig. 7). For instance, in 2012, study S13 witnessed the recognition of symbols using Self-Organizing Map neural network (SOM-NN). Similarly, since 2015, the advent of hybrid architectures like CNN with Recurrent Neural Network, Multi-Layer perceptron, LSTM, or BLSTM have been seen. The recent trend displays the use of varied CNN architectures like Multi-Column Deep neural network (MDNN) and Squeeze-Extracted Multi-feature CNN (SE-MCNN).

It is a critical observation that ML techniques in selected studies are considered individually or combined with other ML or non-ML techniques. The details of all ML-based studies are illustrated in Fig. 8. There have also been observations of non-ML techniques (refer to Fig. 9). The analysis report depicts significantly the prevalent non-ML methods, namely, elastic matching, Hidden Markov Model, relational grammars, and fuzzy logic, which has been employed in 5%(#3), 4%(#2), 4%(#2), and 4%(#2) of the chosen set studies.

Highlights of RQ1: *Most of the recognition models used for mathematical symbol recognition are based on ML approaches. The popular ML algorithms are SVM, K-nearest neighbor, recurrent neural network, and convolutional neural network.*

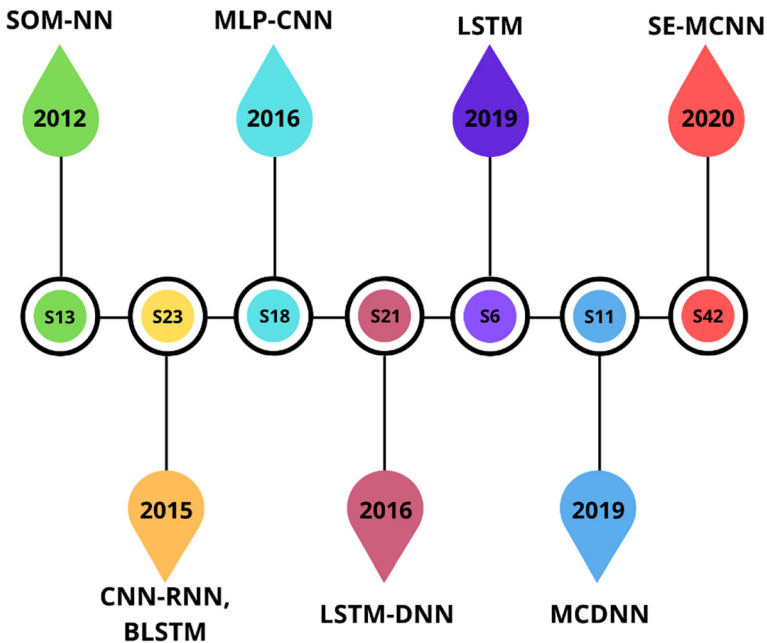


Fig. 7 Deep Learning Architectures over a Timeline

4.2 RQ2 which are corpus mainly used for experimentation in mathematical symbol recognition?

About 24 corpora have been used in the selected studies, but only those corpora are included that are used in more than one study. The most widely used datasets are self-created datasets employed in about 14, i.e., 26% of selected studies. The second most prevalent dataset is the CROHME dataset launched by the CROHME competitions series to recognize handwritten math expressions. This series had a periodic launch of datasets yearly, leading to several versions of the same core dataset. Thus, CROHME has its editions as CROHME 2011 [30], CROHME 2012 [16, 33], CROHME 2013 [15, 53], CROHME 2014 [4, 27], CROHME 2016 [19, 20], and among all primarily used dataset is CROHME 2013, which has been employed in 7, i.e. (13%) of the total selected studies. CROHME 2014 is the second widely investigated dataset, employed in 5, i.e. (9%) of the chosen studies, followed by the dataset CROHME 2016 employed in 4, i.e. (7%) of the collected studies. InfyCDB-1 is investigated in two different studies. Some of the datasets, like CROHME 2012, InfyCDB-3 [48, 49], Ground-truth [29, 47], and Modified National Institute Of Standards And Technology [21, 52], have been maximum investigated in about two studies. The statistical analysis of datasets is presented in a way that the CROHME studies and other recurrent datasets have been segregated and displayed in Fig. 10 in the form of groups. The 1st group elucidates the CROHME studies, whereas the second group depicts the recognized datasets like Modified National Institute Of Standards And Technology, InfyCDB. The self-created category of the dataset is the kind of corpus that has been developed by each researcher as per his requirement and availability of resources. Before the launch CROHME series of competitions, most of the literature used this category where the researchers developed their corpus for implementation.

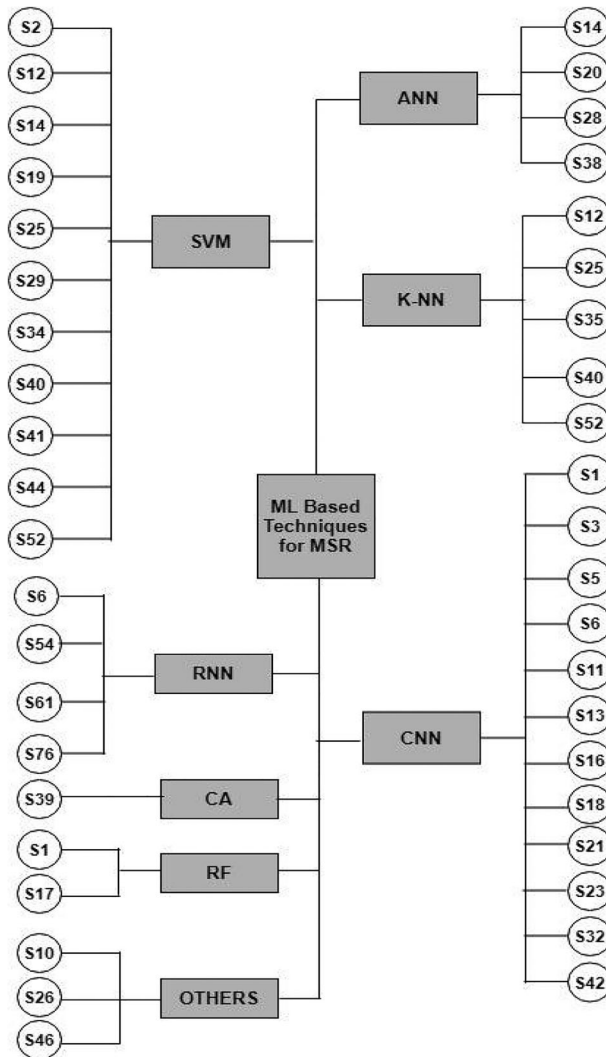


Fig. 8 ML Techniques Identified by SLR Studies

Highlights of RQ2: Majorly, the datasets used for experimentation are self-created datasets used in around 26% of studies. The launch of CROHME datasets has been influential, and the analysis witnesses CROHME to be the second most popular dataset.

4.3 RQ3 what are the majorly used accuracy metrics for mathematical symbol recognition? Identify the notable metric that corresponds to a specific phase in the recognition process?

One of the essential components of the SLR is accuracy metrics, as it is used to underline how efficient a recognition system is to recognize symbols in mathematical symbol recognition. In this SLR, several accuracy measures have been used.

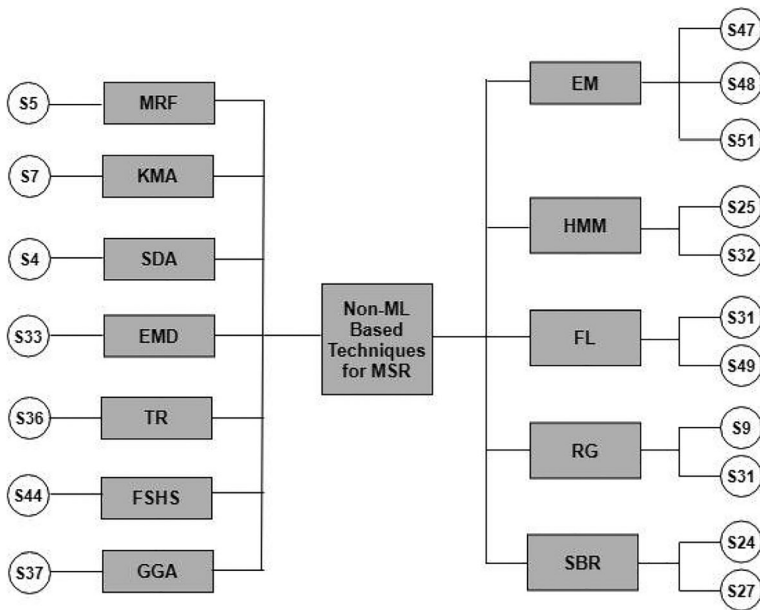


Fig. 9 Non-ML Techniques Identified by SLR Studies

Symbol Recognition Rate (SRR) and Minimum Symbol Error Rate (MSER) has broadly used mathematical symbol recognition studies selected for this SLR. The higher the SRR and lower the Minimum Symbol Error, the more accurate a particular technique is. An SRR is employed in about 36 studies, i.e., 69% of the selected studies and mathematical symbols and expression is implemented in about 18 studies constituting 34% of the selected studies. The third highest accuracy measure is the processing time used in four (7%) of the selected studies. Other accuracy measures are stroke recognition rate, ranking metrics, classification rate, segmentation rate, and convergence rate has been employed in about 2 to 3 studies, constituting overall 4 to 5% of the individual weightage in the total selected studies for this SLR. One of the remarkable factors while choosing the accuracy metrics is that at least two of the chosen studies must witness its usage during the experimentation.

Thus, the effective extract in this research question’s findings is that the notable metric for evaluating recognition results is SRR. Figure 11 illustrates the accuracy metrics, frequency count (green part of the chart), and their weightage (the grey part of the chart), and the

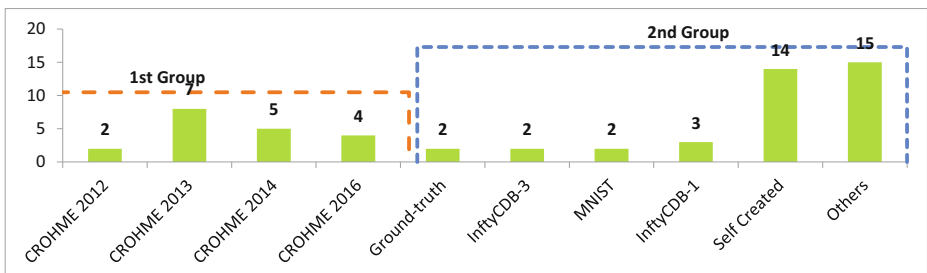


Fig. 10 Corpus employed in the studies selected by SLR



Fig. 11 Accuracy metrics that are employed in different studies in the SLR

evidential studies as the reference. The author has also endeavored to provide brief insight about metric selection criteria and their pros and cons in Table 7.

4.3.1 Insights about validation methodologies for machine learning

Validation methodologies and approaches are the way to determine the error rate in the proposed ML models. ML validation approaches are used to calculate the ML model’s error rate, which is as close to the population’s genuine error rate as possible. The authors have synthesized the validation approaches in all the studies which deployed the ML based symbol recognition model. The analysis for the validation approaches in the studies is depicted in Table 8. The predominant validation techniques implemented in the selected studies are trained/ test split and k-fold cross-validation as depicted in Fig. 12. Few studies didn’t mention any instance of validation methodology, are considered in the category of “Not Applicable”. The train/test split is the most basic strategy. The premise is simple: the data is split into two categories—for instance, 70% for training and 30% for testing the model at random. In k-fold cross-validation, K-1 folds are utilized for training, and the remaining one is used for testing purposes. The benefit of using these methodologies is that all of the data may be used for training and testing.

Highlights of RQ3: *The most popular accuracy metric is the symbol recognition rate deployed in almost 70% of the studies, and other prominent metrics are minimum symbol error rate (34%) and processing time (7%). The well-known validation technique deployed in machine learning-based studies are trained/test split and k-fold cross-validation.*

Table 7 Contemporary analysis of Metric Selection Criteria

Accuracy Metric	Description	When to use	Pros	Cons
Symbol Recognition	The number of successfully detected symbols divided by the total number of symbols constitutes the symbol recognition rate.	This metric should be used when a singular value is needed from a large set of resultant recognized values.	This metric helps to assess the efficacy of models designed for recognizing complex handwritten structures.	Whenever we are dealing with multiple symbols simultaneously, it results in a low recognition rate.
Minimum symbol error rate:	It is the lowest achievable error rate for an outcome classifier.	It is used to set a statistical lower bound on the error that can be achieved for a given classification problem and feature set.	It helps to minimize the probability of symbol decision error.	The approximation involved in the minimum error rate calculation can add a significant amount of uncertainty and ambiguity to the performance analysis.
Mean polynomial distance:	In Euclidean space, mean polynomial distance evaluates the average distance between polynomials.	It is used when a characteristic polynomial of the distance graph matrix is to be evaluated.	It performs well when deployed to datasets that include compact and isolated clusters	If two data vectors have no attributes values in common, they may have a smaller distance than the other pair of data vectors containing the same attribute values. The largest scaled feature will dominate the other features.
Maximum mutual information:	It identifies uncertainty between two random variables. It indicates a large reduction in uncertainty. Low mutual information indicates a small reduction in uncertainty, and zero mutual information means the variables are independent. It is different from the least error rate.	It is deployed in case of uncertain values and their measurement.	One of the quantities measures how much one random variable tells us about another random variable.	It is very challenging and tough to use this metric. Also, without the classifier, maximum mutual information can't be expressed.
Ranking metric	The ranking metrics measure the performance of the Information retrieval system.	Often used to optimize the loss function	The non-differentiability of ranking metrics has sprung up a body of research that endeavors to find differentiable surrogate losses.	It is challenging to optimize directly as ranking metrics are either flat or discontinuous.
Processing Time	In normal situations, it tells us how long it will take to process an application.	Whenever the performance in terms of time is evaluated, the metric processing time comes into the picture. Also, when time is crucial	Effectively measures the performance of the system w.r.t time.	It experiences trade-off issues when mapped along with the space parameters.

Table 7 (continued)

Accuracy Metric	Description	When to use	Pros	Cons
Overlap ratio	Overlap ratio is the ratio of overlapping area to the minimum area of the two connected symbols components.	Whenever we are working on segmentation paths, evaluating overlap ratios becomes crucial.	Evaluating overlap ratio becomes helpful whenever segmentation is to be performed on touching symbol images.	Inherited properties of symbols may add internal challenges to the evaluation and deployment process. After performing segmentation on overlapping areas, when classification algo is deployed, there are fair chances of more false positives and false-negative cases.
Classification rate	The rate of the number of correct predictions to the total number of input samples.	Whenever the performance of the classification model is to be specified, the classification rate is evaluated	Assures the prediction value of the developed classification model.	It fails on classification problems with a skewed class distribution.
Segmentation rate	Rate of the correctly segmented area over ground truth data.	When we measure the performance or quality of the image segmentation model, this metric is evaluated.	It is used to extract objects of interest from images.	It becomes difficult to extract objects with inhomogeneous intensities.

4.4 RQ4 what segmentation/feature extraction techniques are preeminent for mathematical symbol recognition?

The literature displays significant stances of segmentation and feature extraction used in the sub-stages of the recognition process, as mentioned in Tables 9 and 10, respectively. This question targeted the extraction and identification segmentation and feature extraction techniques used in the implementation by most selected studies. There also appeared some studies with no association of segmentation and feature extraction technique. The review group of this SLR decided to map the data extraction of the chosen studies under these two heads to explore another dimension of a striving potential parameter of study.

Highlights of RQ4: *K-NN is found to be a popular segmentation and classification technique. In feature extraction, euclidean distance matching and convolutional neural networks are the most recognized techniques.*

Table 8 Analysis of Validation Techniques

Study no	Validation Technique	Study no	Validation Technique
S1	Train /Test Split	S23	Train /Test Split
S2	Five-fold Cross-Validation	S28	Cross-Validation
S6	Train /Test Split	S29	Five-fold Cross-Validation
S8	Train /Test Split	S34	Train /Test Split
S11	Ten-fold Cross-Validation	S35	Ten-fold Cross-Validation
S12	Five-fold Cross-Validation	S38	Train /Test Split
S13	Train / Test Split	S39	Train /Test Split
S14	Not applicable	S40	Not applicable
S17	Train /Test Split	S41	Five-fold Cross-Validation
S18	Train /Test Split	S42	Train /Test Split
S19	Cross Validation	S45	Train /Test Split
S20	Train /Test Split	S46	Not applicable
S21	Train /Test Split	S52	Three-fold Cross-Validation

4.5 RQ5 what are trend analysis statistics pertinent to the performance of machine learning /non- machine learning techniques with other machine learning and non-machine learning techniques?

This research question is intended to identify which category of technique (ML /non-ML) outperformed the other. The designed approach to answer this question necessitated the need to specify a common ground or a prevalent parameter based on which the conclusions could be stemmed and assimilated. So, several rounds of discussions between the reviewers lead to finalizing a standard accuracy metric for this comparison. However, the ideal stances of efficacious comparisons could have been when both common dataset and concordant accuracy metrics could be brought to use. As the data extraction leads to the insufficient count of such studies, the accuracy metric is taken as a definite and impeccable parameter for comparison. As RQ3 has been suggestive of identifying the SSR as a prominent and frequent accuracy metric kind, the studies' techniques are compared based on SSR, and S30 witnesses the highest

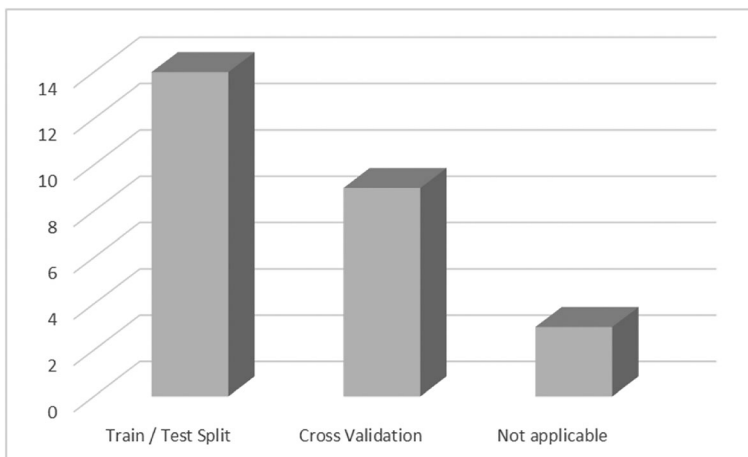
**Fig. 12** Validation metric analysis

Table 9 Segmentation Techniques Identified from SLR

Segmentation Technique	Description	References
K-NN	This classification technique is a non-parametric method where all the computation is deferred until function evaluation. The function is approximated locally, and this technique here is used for image segmentation.	S1, S44
Projection Profile Cutting and Connected Components Labelling Method	The projection profile method is the natural choice for line segmentation of Grayscale images, and by finding the maximum projections values, gaps between the text lines can be found. It can be calculated by summing the pixel values along with the horizontal directions of the document image. The overlapping character segmentation problem is solved by the connected components labeling method. The recognition complexity can be enlarged as this separates simple characters by finding out constituent glyphs.	S12, S46
SOM neural network	This primary image segmentation method is used to reduce dimensionality. Moreover, it is a type of artificial neural network, and it helps in the representation of training samples such that this representation is discrete and low-dimensional.	S13
Contour Tracing Algorithm and Corner Detection Algorithm	Contour Tracing Algorithm, a kind of segmentation technique, is used in the digital region, and it helps in the identification of boundary pixels. It is also called boundary tracing. The corner Detection Algorithm helps in identifying a corner point having low self-similarity.	S15
Simple Linear Iterative Clustering	It generates uniform and compact super-pixels by clustering pixels in image plane space and 5D color. High computation power is not required by these pixels.	S16
Fuzzy Logic	To separate the touching characters, fuzzy logic helps identify cut-point by selecting a row, column, or matrix diagonal element. It is deployed as an unsupervised segmentation part.	S20
Recursive Algorithm	The smaller input values are given to the algorithm, and simple operations are performed on the results obtained. The algorithm calls itself such that solvable solutions should be found for smaller input values.	S22
Candidate Character Lattice Method	To automatically identify the valuable words, this technique is used. The gate vectors are used in this that helps in controlling information flow by using words.	S27
HMM	It has a Markov process having unobserved or hidden states and another dependent process on the Markov process, which yields optimum segmentation. In contrast.	S30
K-means algorithm	This algorithm helps figure out K groups in the unlabeled data, where K represents the number of groups.	S32
Global Segmentation and Local Segmentation	Global Segmentation uses primary labeling on the text lines to separate isolated formulas, whereas Local Segmentation is performed on text lines to separate embedded formulas.	S49
Improved Connected Domain Method	This algorithm can complete the calculation in small time complexity and can improve the running speed	S50

accuracy value for SSR, which used BLSTM-RNN as a recognition method. This concludes that the ML technique outperforms the non-ML recognition models in terms of accuracy metrics, as shown in Table 11.

Table 10 Features Extraction Techniques from SLR

Feature Techniques	Description	References
Functional Approximation	This technique targets a function that closely matches the well-defined class in a task-specific way. It is here for the purpose of feature extraction from collected datasets.	S2, S3
Graph rescoring	This is known as the trigram syntax model to the symbol graph. This model helps in probability computation for each symbol-relation pair by using the preceding two pairs of symbol-relation.	S4
Random Forest Classifier	It uses several decision tree classifiers on different dataset subsamples and uses the average of these different results to predict accuracy and control over-fitting.	S5
Position relation between essential elements	Extraction of features is performed and features like a tail point, middle point, and head point.	S7
Elastic Matching	A 2D optimization problem that specifies corresponding pixels between subjected images.	S10, S43
Wavelet and zoning techniques	To extract discriminant features from a binary image, the c-WZ method combines wavelet techniques and zoning techniques. The feature vector is made that helps in the usage of geometrical and statistical image characteristics.	S12
Euclidean distance	It is a line segment length between two points, and these points are considered in Euclidean space.	S13,S19,S35
Speeded Up Robust Features, Census Transform Histogram, GIST descriptor, Pyramid Histogram Of Oriented Gradients	Speeded Up Robust Features is used for object recognition and feature extraction and is a patented local feature detector and descriptor. CENTRIST is a descriptor used to extract features in topological objects. GIST is used to develop a low dimensional representation of the scene, and Pyramid Histogram Of Oriented Gradients uses image edge lines and makes a fuzzy histogram of gradient directions.	S14
Histogram Of Oriented Gradient and Linear Binary Pattern	Histogram Of Oriented Gradient uses the image localized portions to count gradient orientation occurrences. Linear Binary Pattern uses pixel intensities binary comparison between the center pixel and its eight surroundings.	S16
Shape Context, Chain Code Histogram	Shape Context is used for the description of shapes and similarities of the object. Chain Code Histogram is used to group objects that look similar.	S17, S28
CNN	It is a specific way of artificial neural network that uses perceptron's, and CNN is used for cognitive tasks like image processing and natural language processing	S18, S43
Linear Normalization, Pseudo-2D Bi-Moment Normalization (P2DBMN), Line Density Projection Interpolation	Linear Normalization is the process of bringing the image to the normal range to sense and use in linear data types. P2DBMN: On partitioning the input image into strips, the 1D bi-moment method on the strip's	S21, S23

Table 10 (continued)

Feature Techniques	Description	References
	projection is the estimation method of each strip's 1D coordinate mapping function. Line Density Projection Interpolation: the line density map is partitioned into three horizontal/vertical stripes. Each strip's line density projection gives a 1D coordinate mapping function to avoid the heavy computation of line density blurring.	
Gaussian filter	While using step function input and maintaining minimum fall and rise time, no overshoot filter is used.	S22
Least-square fitting	By minimizing the sum of the squares of the offsets of the points from the curve for finding the best-fitting curve to a given set of points is termed as Least-squares fitting	S25
Multidirectional gradient feature extraction	It represents the map of the gradient with directions in the original image	S26
Offline and Online Feature Extractors	These feature extractors extract features using SVM 1(Online classifier) and SVM 2(Offline classifier)	S29
Parse tree extraction	A parse tree feature space is represented as an n-dimensional predefined feature extraction vector. The two parse trees' similarity is computed by comparing the infinite dimension of the feature vector.	S31
SOM-neural network	This method is used to reduce dimensionality. It has an advantage over conventional feature extraction methods. Moreover, it is a type of ANN, and it helps in the representation of training samples such that this representation is discrete and low-dimensional.	S13,S33
Color-driven reconstruction	Color-driven reconstruction extract features like Hue, saturation, brightness at the beginning of the stroke, and width.	S36

Highlights of RQ5: *On analyzing the accuracy metrics, ML technique: BLSTM-RNN (used in S30) has outperformed other ML and Non-ML methods. Among Non-ML techniques, Elastic Matching (used in S48) achieves the best performance by yielding a symbol recognition rate of 99.70%.*

4.6 RQ6 is the size of the dataset a factor influencing the accuracy rate of mathematical symbol recognition?

The necessary intent for the formulation of this research question is to examine whether the dataset's size is a dependable parameter for affecting the accuracy rate of mathematical symbol recognition. To formulate the answers to this research question, the data extraction has been done in the context of two dataset size parameters: dataset classes and the number of symbols included in the dataset. The accuracy values extracted from the studies are mapped corresponding values of dataset parameters, and the mapped charts are presented in Figs. 13 and 14. According to the mapped charts, the values are arranged on the designed metric-size scale

Table 11 Accuracy Metric Analysis

Studies	Technique used	ML/Non-ML	Accuracy Metric	Value %
S1	Random Forests	ML	Symbol Recognition rate (SRR)	88.85%
S2	SVM	ML	The minimum Symbol Error rate	7.28%.
S3	Clustering K-means	ML	Mean Polynomial Distance	75.55%
S4	Symbol Detecting Algorithm	Non-ML	Maximum Mutual Information (MMI)	96.94%
S5	Markov Random Field	Non-ML	SRR, Stroke recognition rate	95.80%, 97.80%
S6	Long Sort-Term Memory (Long Sort-Term Memory)	ML	Ranking metric	0.45
S7	Kohn-Munkres algorithm	Non-ML	SRR	90.52%
S8	BLSTM-RNN	ML	SRR	84.10%
S9	Relational grammars	Non-ML	SRR	96.20%
S10	Adaptive Resonance Theory	Non-ML	Minimum Symbol Error rate	0.97%
S11	CNN	ML	SRR	92.02%
S12	SVM, K-NN	ML	SRR, Ranking metric	96.54%, 1.66
S13	Field Programmable Gate Array embedded system	ML	Processing Time	85%
S14	SVM, ANN	ML	SRR	99%
S15	Corner Detection Algorithm	Non-ML	Overlap ratio	
S16	LeNet and SqueezeNet	ML	SRR	90%
S17	Random Forests	ML	SRR	97.95%
S18	CNN	ML	Minimum Symbol Error rate, Classification rate	12.28%, 87.72%
S19	SVM	ML	SRR	97%
S20	Backpropagation algorithm, ANN	ML	Segmentation rate	93.6%
S21	CNN, BLSTM-RNN	ML	SRR	91.28%
S22	Contouring Approach		Segmentation rate	76.00%
S23	CNN, BLSTM-RNN	ML	SRR	91.28%
S24	Stroke based recognition	Non-ML	SRR	98.38%
S25	SVM, K-NN, Hidden Markov Model (HMM)	Both	Minimum Symbol Error rate	97.40%
S26	Multidirectional feature fusion decision discriminant algorithm	Non-ML	SRR	93.12%
S27	Stroke based recognition	Non-ML	Segmentation rate	97.10%
S28	Multi-Layer Perceptron neural network	ML	SRR	86%
S29	SVM	ML	SRR	5% (online recognizer) 22% (offline recognizer)
S30	BLSTM-RNN	ML	SRR	99.72%
S31	Fuzzy logic, Relational grammars	Non-ML	SRR, Stroke recognition rate	85.99%, 71.73%
S32	Hidden Markov Model (HMM)	Non-ML	SRR	82.90%
S33	Earth's Movers Distance	Non-ML	SRR	98.10%
S34	SVM	ML	SRR	81.55%
S35	K-NN	ML	Minimum Symbol Error rate	97.50%
S36	Trajectory Recovery	Non-ML	SRR	84.85%

Table 11 (continued)

Studies	Technique used	ML/Non-ML	Accuracy Metric	Value %
S37	Graph generation algorithm	Non-ML	Minimum Symbol Error rate	80.90%
S38	Niche genetic algorithm-BP neural network model	ML	SRR	90.60%
S39	ISOETRP	ML	SRR	95.60%
S40	clustering Algorithm SVM, K-NN	ML	Minimum Symbol Error rate	
S41	SVM	ML	SRR, Minimum Symbol Error rate	97.70%
S42	CNN	ML	SRR	92.44%
S43	CNN	ML	SRR	
S44	4-step homotopy strategy	Non-ML	SRR	99.45%
S45	SVM	ML	SRR	89.87%
S46	coarse classification-to-fine recognition strategy	ML	SRR	97.81%
S47	Elastic Matching	Non-ML	SRR	97.60%
S48	Elastic Matching	Non-ML	SRR	99.70%
S49	Fuzzy logic	Non-ML	SRR	94.65%
S50	Improved Connected Domain Algorithm	Non-ML	SR	83.20%
S51	Elastic Matching	ML	SRR	95.50%
S52	SVM, K-NN	ML	Classification rate	92%

(Fig. 15), which depicts the accuracy ranges concerning the size of symbol classes. The data analysis helped to conclude the following points:

- The symbol class of size range from 0 to 100 (as shown in Fig. 15 gives the accuracy range up to 89.97%, leading to a critical point that the accuracy range is from 0 to 100 can reach the accuracy values less than 90%.
- There are highlighted entries of two accuracy values that are misleading the results that are accuracy values produced by the studies S17 and S30. The authors consider these two results of accuracy values to be a possible outcome of overfitting stances.
- The symbol class of size range from 100 to 300 presented on the accuracy metric size scale is definite for producing the accuracy values equal to or more than 90% and lies less than 97%.
- The symbol class ranging in size 300–400 outputs the accuracy values of more than 97%. So, the accuracy metric-size scale is entirely suggestive that dataset parameters size can be a dependent factor for determining the accuracy values.
- The authors also believe that the dataset cannot be the only parameter for influencing the accuracy values. The accuracy results can depend on other factors like recognition technique employed, kind of dataset used, accuracy metric type, specific pre-processing, segmentation, and feature extraction methods. These all can also be contributing factors for determining the accuracy range of a recognition process, as the size of the dataset has been co-related with the accuracy obtained and presented in Fig. 15.

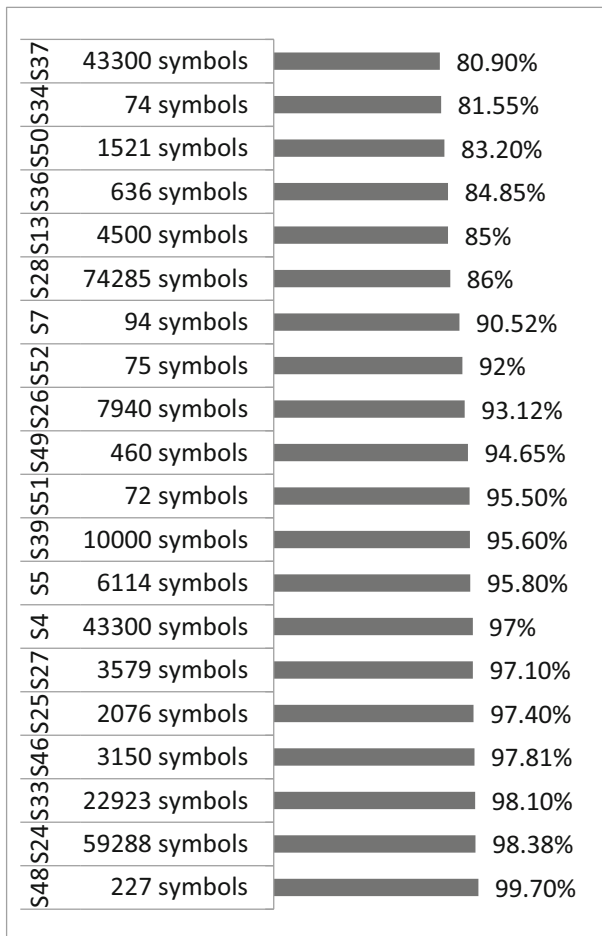


Fig. 13 Accuracy based on total symbols

Highlights of RQ6: Accuracy metric when mapped to dataset size, the scale (Fig. 15) depicts that when the dataset has more than 300 symbol classes, the recognition process is likely to achieve accuracy of 97% or above. The two exceptional studies (S17, S30) have been noted to provide high accuracy even with fewer symbol classes.

4.7 RQ7 what are the predominant publication channels for research communications in mathematical symbol recognition?

Knowing about the articles' primary sources related to mathematical symbol recognition is the idea behind this research question. International Journal on Document Analysis and Recognition (IJ DAR) has published many studies of mathematical symbol recognition. It is the leading journal among all that have analyzed in this SLR, 5% of different studies in the SLR published in this journal, and International Conference on Document Analysis and Recognition (ICDAR) is the leading conference among other conferences that have been analyzed in our SLR, 23%

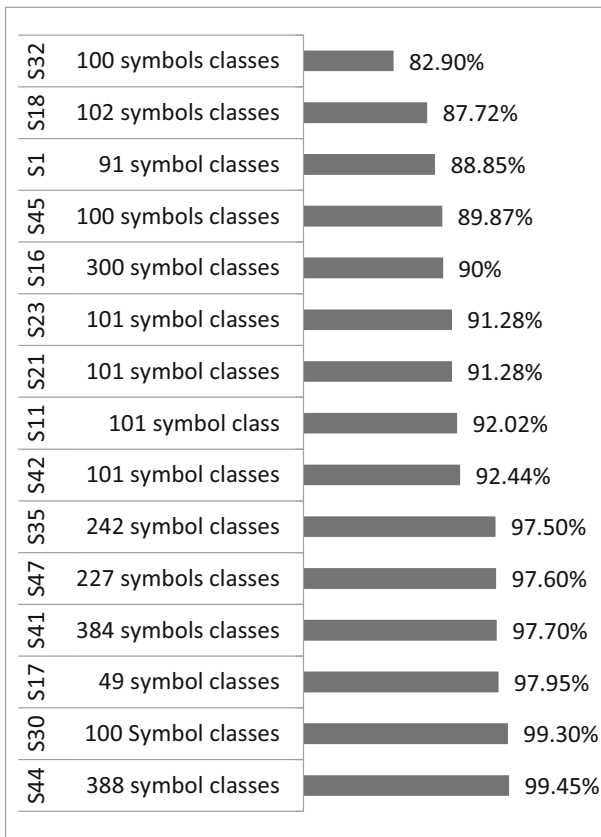


Fig. 14 Accuracy based on Symbol Classes

of different studies in our SLR published in this conference. International Conference on Frontiers in Handwriting Recognition (ICFHR) published 7% of other studies in our SLR and International Conference on Pattern Recognition (ICPR) 7% of different studies in our SLR are

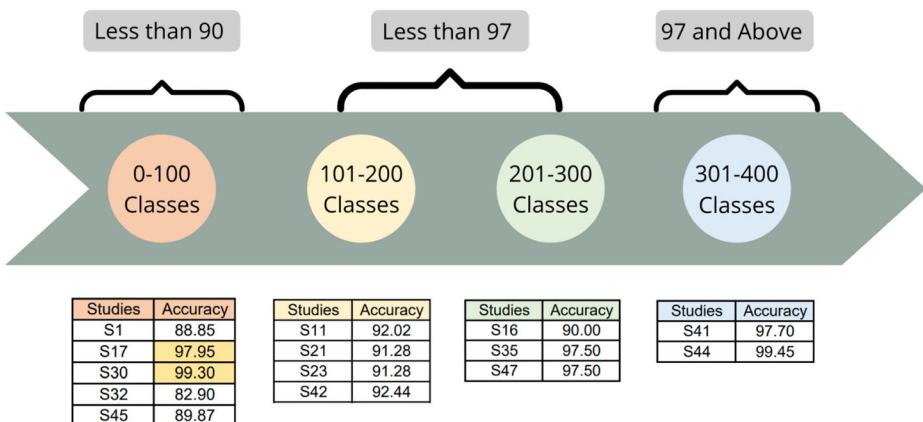


Fig. 15 Accuracy Metric-Size Scale

the following major conferences publication sources. The sources which have at least two publications studies in our SLR are considered in this research question.

Highlights of RQ7: *The predominant channel for publication is a renowned conference ICDAR which conducts the CROHME series of competitions, and around 23% of reviewed studies are part of this conference proceedings.*

4.7.1 Extending discussions on problems encountered during math symbol recognition

The following are the five main problems associated with recognizing math notations, symbols, and expressions. Four of the issues are given by [78], and our analysis adds one more problem.

- **Expression Extraction/ Expression Detection:** It's tough to extract or recognize a mathematical expression on the same line as ordinary text. Tracing an exact phrase surrounded by words and extracting it from a text-rich document confuses and complicates the recognition operation. Furthermore, detecting expressions hidden in large amounts of textual material is difficult.
- **Symbol Extraction/Symbol Recognition:** There are two types of representations involved. One is vector-based representations, and the other is raster-scanned representations. In the case of the former, we have PDFs; numerous symbol labels and locations can be recovered, but it requires handling of some exceptional cases like the one is mentioned by [5], a root symbol is not a single glyph (a connected component). It comprises a typeset (radial symbol) with an upper horizontal bar represented differently from a radial sign. In the latter case, where raster image data drawn with pen strokes is available. The complicated task is to detect and identify the symbol's location. There are hundreds of alphanumeric and mathematical symbols used, many so similar in appearance that some use of context is necessary for the point of disambiguation. (e.g., o O 0 add reference) .
- **Layout Analysis:** It is challenging to perform a detailed analysis of the spatial relationships between distinct symbols of varied scales embedded in the math expressions. When these spatial relationships among the symbols are represented using tree structures, we term them as Symbol Layout Trees. These symbol layout trees also represent contextual information like that present in LATEX MEs. They specify the groups that belong to horizontally adjacent symbols, and they share a baseline or writing line, along with superscript, below, subscript, above, and having containment relationships. In other words, it can be stated that layout analysis is all about a more in-depth analysis of the spatial relationships among the math symbols using symbol layout trees. Layout analysis is performed by many different techniques, and some of these are based on baseline structure trees (BST) [66, 79, 80] projection-profile cutting [18, 54], and graph rewriting [36]. One of the open-source structural analyzers is the Diagram Recognition Application for Computer Understanding of Large Algebraic Expressions (DRACULAE) and is very popular [13, 57, 70] which implements the BST method along with various optimizations and improvements.
- **Semantic Analysis:** This semantic analysis step focuses on understanding the semantics or meaning of the recognized symbols from the segmented expressions. It mainly involves the creation of a semiotic analysis tree that is outputted from the semantic understanding step. It holds the mathematical meaning of the input formula, including the mathematically syntactic

relationship among characters and symbols. The layout analysis and semantic analysis steps have been proposed by [12]. They have also given the basic idea of the system architecture for layout recognition and semantic understanding of mathematical formulae.

Apart from the problems mentioned above, there is a colossal challenge encountered when it comes to comparing different recognition techniques for concluding the best one out of all. In the past, many of the studies have introduced various schemes for recognizing mathematical symbols and MEs. It has been found that mainly, each study has used a distinct method for evaluating the performance accuracy of the recognition systems. Thereby making it a challenging task to compare several approaches to this problem accurately. From the literature, it also has been concluded that several metrics have been recommended to evaluate the MEs recognition systems performance. Some metrics are the symbol recognition rate [10] and symbol segmentation rate [65] that can be calculated if the ground truth is available. Above mentioned metrics take into consideration only a specific part of the recognition problem. Thereafter, the need of standardization of dataset, and a standard accuracy metric has been realized so that the task of comparison could be accomplished more effectively.

5 Summary of findings

The results of the SLR related to mathematical symbol recognition are what we offered, and we chose a total of 52 studies. We extracted studies until 2021, and the sources from where we extracted studies are Scopus, IEEE Xplore, Springer, Google scholar. To highlight the contribution of studies, we have presented the publication year and the number of studies published in that range in our SLR shown in Fig. 16. It is worth noting that from 2006 to 2016, there has been a significant improvement in the publications related to mathematical symbol recognition compared to previous years. As in the inclusion-exclusion criteria, we only focused on mathematical symbol recognition instead of the whole expression. By keeping this in mind, we formed inclusion-exclusion questions. After applying these criteria, we filtered 75 studies, and after this Quality assessment test is applied to the remaining 75 studies. The studies that do not fulfill the quality assessment requirements have been discarded. We got

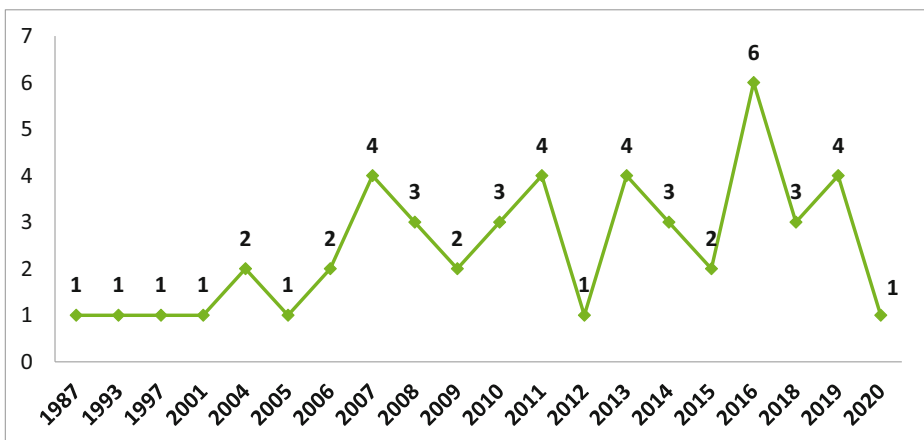


Fig. 16 Studies Distribution from 1987 to 2021

only 52 final selected studies for our SLR. The chosen 52 studies answered the questions in our SLR significantly.

The study analyzed the pattern of research trends in the domain of handwritten mathematical symbol recognition, and it has been observed that the trend line is focused more on ML techniques. Amidst the ML techniques, SVM is the most well-known technique. K-Nearest Neighbour and CNN as the second most frequent technique used. To sum the observation, the authors have endeavored to summarize in Table 12.

6 Conclusion and future scope

In this paper, we have covered all the studies until 2021 related to mathematical symbol recognition. A total of 52 studies we have selected in our SLR to evaluate and analyze critical information for users. A substantial amount of work has been done in mathematical symbol recognition, and a systematic review of it is a must need and this idea behind our SLR. In results and comparisons, we found that SVM is the most used ML technique and elastic matching is the most used non-ML technique. The most employed accuracy parameter is SSR and MSE in studies selected in our SLR. Comparing techniques used SSR and feature set or feature extraction is the broadly used metrics in studies. In terms of dataset/corpus, CROHME is the most used corpus in our SLR studies. Among CROHME versions, CROHME 2013 has been used widely in our SLR. This organized survey study has also extracted the uncovered and less focused potential dimensions of this domain. The reviewers have extracted the following points for future research in this direction:

- Ambiguity and the 2D structure of MEs have always remained a challenge in the task of recognition. Moreover, the literature witnesses a decent scope and potential that involves more complicated MEs that work on high-order exponentials.

Table 12 Research Questions Major Extracts

ID	Research Questions	Major Extracts
RQ1	Which ML/non-ML techniques are employed in mathematical symbol recognition?	Several recognition techniques have been analyzed, and SVM has been found as the most implemented recognition technique.
RQ2	Which are corpus mainly used for experimentation in Mathematical symbol recognition?	CROHME dataset and self-written datasets.
RQ3	What are the majorly used accuracy metrics for mathematical symbol recognition?	Symbol Recognition Rate
RQ4	What are segmentation/feature extraction techniques preeminent for mathematical symbol recognition?	Major Segmentation technique: K-NN, Minimum Distance Classifier Primary Feature Extraction technique: Euclidean Distance
RQ5	What are trend analysis statistics pertinent to the performance of ML techniques with other ML and non-ML techniques?	ML techniques (57% of studies) outperformed the non-ML models (43% studies).
RQ6	Is the size of the dataset a factor influencing the accuracy rate of mathematical symbol recognition?	Yes, the size of the dataset can be a dependent factor for determining the accuracy values.
RQ7	What are the predominant publication channels for research communications in mathematical symbol recognition?	Predominant journal: IJDAR Predominant Conference: ICDAR

- On analyzing the several representation models implemented in the recognition process, it has been found that there is a need for a more efficient representation model or structure that could contain information for both symbol segmentation and stroke segmentation.
- The main challenge for the symbol segmentation sub-step is a lack of clear rules to group the strokes belonging to a single symbol embedded in the whole math expression.
- One of the constraints observed during the review process is defining the metric accuracy scale (Fig. 15). There is a limitation while describing the relationship between accuracy metric and dataset size. That is because of the absence of standard publically available datasets and standard accuracy metrics. Until we don't get common grounds in varied recognition approaches, it is hard to figure out the direct relationships between the necessary comparison factors. Performing comparative analysis by narrowing the field and choosing specific intersecting factors for comparison is not a good idea.
- The datasets available in this domain of handwritten mathematical symbols and expressions needed more semantic-based handwritten symbols and expressions involving attributes like age, gender, etc. An improved recognizer could be built for advancement in this domain.
- It can also be observed that multi-modal encoder-decoder systems that have been deployed for recognition can have a fair scope for offline to online reduction resulting in a single recognizer for both types of math expressions. Exploring the trend of multi-attention and multi-modal systems and working on its enhancement can be a said challenge in itself.
- Ambiguity problems arising at the pre-processing step are challenging as scaling may lead to distortion and the shape formation of specific signs and symbols. The real meaning could get distorted in this process.

Appendix A

Table 13 Table of Acronyms

Abbreviation	Full form	Abbreviation	Full form
ANN	Artificial Neural Network	ML	Machine Learning
BLSTM	Bidirectional Long Short Term Memory	MLP	Multi-Layer Perceptron
ConRate	Convergence Rate	MSER	Minimum Symbol Error Rate
CR	Classification Rate	OR	Overlap ratio
CROHME	Competition On Recognition Of Online Handwritten Mathematical Expression	OSR	Over-Segmentation Rate
EM	Elastic Matching	RA	Recursive Algorithm
EMD	Earth Mover's Distance	RF	Random Forests
FL	Fuzzy Logic	RM	Ranking Metrics
FL	Fuzzy Logic	SDA	Symbol Decoding Algorithm
GGA	Graph Generation Algorithm	SR	Segmentation Rate
HMM	Hidden Markov Model	SRR	Symbol Recognition Rate
KMA	Kohn-Munkres Algorithm	StR	Stroke Recognition Rate
KNN	K-Nearest Neighbour	SVM	Support Vector Machine
LSTM	Long Sort-Term Memory	TR	Trajectory Recovery
ME	Mathematical Expressions	USR	Under Segmentation Rate

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Ali I, Mahjoub M (2018) Dynamic random forest for the recognition of arabic handwritten mathematical symbols with a novel set of features. *Int Arab J Inf Technol* 15(3A special issue):565–575
2. Álvaro F, Sánchez JA (2010) Comparing several techniques for offline recognition of printed mathematical symbols. International conference on pattern recognition, 1953–1956. <https://doi.org/10.1109/ICPR.2010.481>
3. Alvaro F, Sanchez JA, Benedi JM (2013) Classification of on-line mathematical symbols with hybrid features and recurrent neural networks. Proceedings of the international conference on document analysis and recognition, ICDAR, 1012–1016. <https://doi.org/10.1109/ICDAR.2013.203>
4. Alvaro F, Sanchez JA, Benedi JM, Sánchez J-A, Benedi J-M (2014). Offline features for classifying handwritten math symbols with recurrent neural networks. 22nd international conference on pattern recognition, 2944–2949. <https://doi.org/10.1109/ICPR.2014.507>
5. Baker JB, Sexton AP, Sorge V (2010) Faithful mathematical formula recognition from PDF documents. *9th IAPR international workshop on document analysis systems*, 485–492. <https://doi.org/10.1145/1815330.1815393>
6. Bouvett E, Casha O, Grech I, Cutajar M, Gatt E, Micallef J (2012) An FPGA embedded system architecture for handwritten symbol recognition. *Proceedings of the Mediterranean Electrotechnical conference - MELECON*, 653–656. <https://doi.org/10.1109/MELCON.2012.6196516>
7. Chajri Y, Maarir A, Bouikhalene B (2016) A comparative study of handwritten mathematical symbols recognition. Thirteenth international conference on computer graphics, imaging and visualization, 448–451. <https://doi.org/10.1109/CGiV.2016.92>
8. Chan KF, Yeung DY (2000a) Mathematical expression recognition: A survey. *Int J Doc Anal Recognit (IJDR)* 3(1):3–15. <https://doi.org/10.1007/PL00013549>
9. Chan K, Yeung D (2000b) Novel application of on-line mathematical expression recognition technology. Proceedings of sixth international conference on document analysis and recognition (ICDAR). IEEE, 200, 774–778
10. Chan K-F, Yeung DY (2001) Error detection, error correction and performance evaluation in on-line mathematical expression recognition. *Pattern Recogn* 34(8):1671–1684. [https://doi.org/10.1016/S0031-3203\(00\)00102-3](https://doi.org/10.1016/S0031-3203(00)00102-3)
11. Char BW, Watt SM (2007) Representing and characterizing handwritten mathematical symbols through succinct functional approximation. Proceedings of the international conference on document analysis and recognition, ICDAR, 2, 1198–1202. <https://doi.org/10.1109/ICDAR.2007.4377105>
12. Chen Y, Okada M (2001) Structural analysis and semantic understanding for offline mathematical expressions. *Int J Pattern Recognit Artif Intell* 15(EC06):967–987. <https://doi.org/10.1142/S021800140100126X>
13. Clark R, Kung Q, Van Wyk A (2013) System for the recognition of online handwritten mathematical expressions. *Eurocon 2013:2029–2035*. <https://doi.org/10.1016/j.ympev.2006.04.014>
14. Dai Nguyen H, Duc Le A, Nakagawa M (2016) Recognition of online handwritten math symbols using deep neural networks. *IEICE Trans Inf Syst* E99.D:3110–3118. <https://doi.org/10.1587/transinf.2016EDP7102>
15. Kenny Davila; Richard Zanibbi. (2018). Visual search engine for handwritten and typeset math in lecture videos and LATEX notes. 2018 16th international conference on Frontiers in handwriting recognition, ICFHR, 2018-Augus, 50–55. <https://doi.org/10.1109/ICFHR-2018.2018.00018>
16. Davila K, Ludi S, Zanibbi R (2014) Using off-line features and synthetic data for on-line handwritten math symbol recognition. Fourteenth international conference on Frontiers in handwriting recognition, 323–328. <https://doi.org/10.1109/ICFHR.2014.61>
17. Dimitriadis YA, Coronado JL, Moreno CG, Izquierdo JMC (1993) On-line handwritten symbol recognition, using an ART based neural network hierarchy. IEEE international conference on neural networks - conference proceedings, 1993-Janua, 944–949. <https://doi.org/10.1109/ICNN.1993.298684>
18. Drsouza L, Mascarenhas M (2018) Offline handwritten mathematical expression recognition using convolutional neural network. International conference on information, communication, engineering and technology, 1–3. <https://doi.org/10.1109/ICICET.2018.8533789>

19. Fang D, Zhang C (2020) Multi-feature learning by joint training for handwritten formula symbol recognition. *IEEE Access* 8(2):48101–48109. <https://doi.org/10.1109/ACCESS.2020.2979346>
20. Fang D, Feng G, Yang H (2019) Gabor features assist semantic feature learning for handwritten formula symbol recognition. 2019 IEEE 9th international conference on electronics information and emergency communication, 230–233. <https://doi.org/10.1109/ICEIEC.2019.8784656>
21. Farulla GA, Armano T, Capietto A, Murru N, Rossini R (2016) Artificial neural networks and fuzzy logic for recognizing alphabet characters and mathematical symbols. *International conference on computers helping people with special needs*, 7–14. https://doi.org/10.1007/978-3-319-41264-1_1
22. Firdaus SA, Vaidehi K (2020) Handwritten mathematical symbol recognition using machine learning techniques: review. *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, 658–671. https://doi.org/10.1007/978-3-030-24318-0_75
23. Garain U, Chaudhuri BB, Ghosh RP (2004) A multiple-classifier system for recognition of printed mathematical symbols. *Proceedings of the 17th international conference on pattern recognition*, 1, 380–383. <https://doi.org/10.1109/ICPR.2004.1334131>
24. Golubitsky O, Watt SM (2009) Confidence measures in recognizing handwritten mathematical symbols. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 5625 LNAI, 460–466. https://doi.org/10.1007/978-3-642-02614-0_36
25. Golubitsky O, Watt SM (2010) Distance-based classification of handwritten symbols. *Int J Doc Anal Recognit* 13(2):133–146. <https://doi.org/10.1007/s10032-009-0107-7>
26. Green BN, Johnson CD, Adams A (2006) Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Chiropr Med* 5(3):101–117
27. Guan SK, Moh M, Moh TS (2019) Context-based multi-stage offline handwritten mathematical symbol recognition using deep learning. *International conference on high performance computing and simulation, HPCS 2019*, 185–192. <https://doi.org/10.1109/HPCS48598.2019.9188180>
28. Hu R, Watt SM (2013) Identifying features via homotopy on handwritten mathematical symbols. *Proceedings - 15th international symposium on symbolic and numeric algorithms for scientific computing, SYNASC 2013, 1*, 61–67. <https://doi.org/10.1109/SYNASC.2013.15>
29. 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. *International conference on document analysis and recognition*, 457–462. <https://doi.org/10.1109/ICDAR.2011.98>
30. Hu L, Hart K, Pospelov R, Zanibbi R (2012) Baseline extraction-driven parsing of handwritten mathematical expressions. *21st international conference on pattern recognition*, 326–330
31. Jakjoud W (2009) Representation, handling and recognition of mathematical objects: state of the art. *Proceedings of the 2009 3rd international conference on research challenges in information science, RCIS 2009*, 427–438. <https://doi.org/10.1109/RCIS.2009.5089307>
32. Jakjoud W, Lazrek A (2011) Segmentation method of offline mathematical symbols. *International conference on multimedia computing and systems*, 1–7. <https://doi.org/10.1109/ICMCS.2011.5945634>
33. Jimenez ND, Nguyen L (2013) Recognition of handwritten mathematical symbols with PHOG features
34. Julca-Aguilar F, Hirata NST, Viard-Gaudin C, Mouchere H, Medjkoune S (2014) Mathematical symbol hypothesis recognition with rejection option. *2014 14th international conference on Frontiers in handwriting recognition*, 500–505. <https://doi.org/10.1109/ICFHR.2014.90>
35. Kacem A, Belaid A, Ben Ahmed M (2001) Automatic extraction of printed mathematical formulas using fuzzy logic and propagation of context. *Int J Doc Anal Recognit* 4(2):97–108. <https://doi.org/10.1007/s100320100064>
36. Kanahori T, Tabata K, Cong W, Tamari F, Suzuki M (2000) On-line recognition of mathematical expressions using automatic rewriting method. *International conference on multimodal interfaces*, 394–401. https://doi.org/10.1007/3-540-40063-x_52
37. Keshari B, Watt SM (2007) Hybrid mathematical symbol recognition using support vector machines. *Ninth international conference on document analysis and recognition*, 2, 859–863. <https://doi.org/10.1109/ICDAR.2007.4377037>
38. Keshari B, Watt SM (2008) Online mathematical symbol recognition using svms with features from functional approximation. *Proc. Mathematical User-Interfaces Workshop*
39. Kitchenham B (2004) Procedures for performing systematic reviews. *Ann Saudi Med* 33(1):79–83. <https://doi.org/10.5144/0256-4947.2017.79>
40. Kurtzberg JM (1987) Feature analysis for symbol recognition by elastic matching. *IBM J Res Dev* 31(1): 91–95. <https://doi.org/10.1147/rd.311.0091>
41. Labahn SMG (2013) A new approach for recognizing handwritten mathematics using relational grammars and fuzzy sets. *Int J Doc Anal Recognit* 16(2):139–163. <https://doi.org/10.1007/s10032-012-0184-x>
42. Lee H, Lee M (1994) Understanding mathematical expressions using procedure-oriented transformation. *Pattern Recogn Lett* 27(3):447–457

43. Liu CL, Yin F, Wang DH, Wang QF (2013) Online and offline handwritten Chinese character recognition: benchmarking on new databases. *Pattern Recogn* 46(1):155–162. <https://doi.org/10.1016/j.patcog.2012.06.021>
44. Luo ZX, Shi Y, Soong FK (2008) Symbol graph based discriminative training and rescoring for improved math symbol recognition. ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings, 1953–1956. <https://doi.org/10.1109/ICASSP.2008.4518019>
45. MacLean S, Labahn G (2015) A Bayesian model for recognizing handwritten mathematical expressions. *Pattern Recogn* 48(8):2433–2445. <https://doi.org/10.1016/j.patcog.2015.02.017>
46. Mahdavi M, Condon M, Davila K, Zanibbi R (2019) LPGA: line-of-sight parsing with graph-based attention for math formula recognition. International conference on document analysis and recognition, 647–654. <https://doi.org/10.1109/ICDAR.2019.00109>
47. Malon C, Uchida S, Suzuki M (2006) Support vector machines for mathematical symbol recognition. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 4109 LNCS, 136–144. https://doi.org/10.1007/11815921_14
48. Malon C, Uchida S, Suzuki M (2008) Mathematical symbol recognition with support vector machines. *Pattern Recogn Lett* 29(9):1326–1332. <https://doi.org/10.1016/j.patrec.2008.02.005>
49. Marinai S, Miotti B, Soda G (2011) Using earth mover's distance in the bag-of-visual-words model for mathematical symbol retrieval. In: Proceedings of the international conference on document analysis and recognition, ICDAR, pp 1309–1313. <https://doi.org/10.1109/ICDAR.2011.263>
50. Medjkoune S, Mouchère H, Petitrenaud S, Viard-gaudin C (2011) Handwritten and audio information fusion for mathematical symbol recognition. International conference on document analysis and recognition, 379–383. <https://doi.org/10.1109/ICDAR.2011.84>
51. Mohamed Shaffril HA, Samsuddin SF, Abu Samah A (2021) The ABC of systematic literature review: the basic methodological guidance for beginners. *Qual Quant* 55(4):1319–1346
52. Nazemi A, Tavakolian N, Fitzpatrick D, Fernando CA, Suen CY (2019) Offline handwritten mathematical symbol recognition utilising deep learning In *Computer Vision and Pattern Recognition*
53. Nguyen DH, Le Duc A, Nakagawa M (2015)s Deep neural networks for recognizing online handwritten mathematical symbols. Third IAPR Asian Conference on Pattern Recognition Deep, 121–125. <https://doi.org/10.1109/ACPR.2015.7486478>
54. Okamoto M, Imai H, Takagi K (2001) Performance evaluation of a robust method for mathematical expression recognition. In: Sixth international conference on document analysis and recognition, pp 121–128. <https://doi.org/10.1109/ICDAR.2001.953767>
55. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, ... Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
56. Pathak A, Pakray P, Das R (2019) LSTM neural network based math information retrieval. 2019 2nd international conference on advanced computational and communication paradigms, ICACCP 2019, 1–6. <https://doi.org/10.1109/ICACCP.2019.8882887>
57. Pillay A (2014) Intelligent combination of structural analysis algorithms: application to mathematical expression recognition. Rochester Institute of Technology
58. Průša D, Hlaváč V (2007) Mathematical formulae recognition using 2D grammars. 2017 Ninth international conference on document analysis and recognition, 2, 849–853. <https://doi.org/10.1109/ICDAR.2007.4377035>
59. Ramadhan I, Purnama B, Al Faraby S (2016) Convolutional neural networks applied to handwritten mathematical symbols classification. In 00 (Ed.), *4th international conference on information and communication technology* (pp. 1–4). <https://doi.org/10.1109/ICoICT.2016.7571941>
60. Ramirez-Pina C, Sanchez JS, Valdovinos-Rosas RM, Hernández-Servín JA (2018) A hybrid feature extraction method for offline handwritten math symbol recognition. *Iberoamerican Congress on Pattern Recognition*, 1, 893–901. <https://doi.org/10.1007/978-3-030-13469-3>
61. Rong LL, Li WJ, Wang G (2009) The research of ISOETRP clustering algorithm on optical mathematical symbols recognition. Proceedings - 2009 9th international conference on hybrid intelligent systems, HIS 2009, 3, 433–436. <https://doi.org/10.1109/HIS.2009.301>
62. Sakshi, Kukreja V (2021) A retrospective study on handwritten mathematical symbols and expressions : classification and recognition. *Eng Appl Artif Intell* 103:104292. <https://doi.org/10.1016/j.engappai.2021.104292>
63. Saroui BS, Sorge V (2015) Trajectory recovery and stroke reconstruction of handwritten mathematical symbols. International conference on document analysis and recognition, 1051–1055. <https://doi.org/10.1109/ICDAR.2015.7333922>

64. Shi Y, Li HY, Soong FK (2007) A unified framework for symbol segmentation and recognition of handwritten mathematical expressions. 9th international conference on document analysis and recognition, 2, 854–858. <https://doi.org/10.1109/ICDAR.2007.4377036>
65. Takiguchi Y, Okada M, Miyake Y (2005) A fundamental study of output translation from layout recognition and semantic understanding system for mathematical formulae. Eighth international conference on document analysis and recognition, 745–749. <https://doi.org/10.1109/ICDAR.2005.10>
66. Tapia E, Rojas R (2003) Recognition of on-line handwritten mathematical formulas in the e-chalk system. Seventh international conference on document analysis and recognition, 3, 980–984. <https://doi.org/10.1109/ICDAR.2003.1227805>
67. Tian X-D, Li H-Y, Li X-F, Zhang L-P (2006) Research on symbol recognition for mathematical expressions. First international conference on innovative computing, information and control, 357–360. <https://doi.org/10.1109/iccic.2006.506>
68. Tian XD, Zuo LN, Yang F, Ha MH (2007) An improved method based on gabor feature for mathematical symbol recognition. 2007 international conference on machine learning and cybernetics, 3, 1678–1682. <https://doi.org/10.1109/ICMLC.2007.4370417>
69. Toyozumi K, Yamada N (2004) A study of symbol segmentation method for handwritten mathematical. Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004, 2, 2–5
70. Tran GS, Huynh CK, Le TS, Phan TP, Bui KN (2018) Handwritten mathematical expression recognition using convolutional neural network. 3rd international conference on control, robotics and cybernetics, 15–19. <https://doi.org/10.1109/CRC.2018.00012>
71. Wang C, Mouchère H, Viard-Gaudin C, Jin L (2016a) Combined segmentation and recognition of online handwritten diagrams with high order Markov random field. International conference on Frontiers in handwriting recognition, 252–257. <https://doi.org/10.1109/ICFHR.2016.0056>
72. Wang H, Wang Y, Lu L, Liu J, Li S, Zhang Y (2016b) An improved algorithm for symbol segmentation of mathematical formula images. 16th international symposium on communications and information technologies, ISCIT 2016, 461–464. <https://doi.org/10.1109/ISCIT.2016.7751674>
73. Wang J, Du J, Zhang J (2020) Stroke constrained attention network for online handwritten mathematical expression recognition. Pattern Recogn 119:1–29. <http://arxiv.org/abs/2002.08670>. Accessed 3 Nov 2020
74. Watt SM, Xie X (2005) Prototype pruning by feature extraction for handwritten mathematical symbol recognition. <http://www.csd.uwo.ca/~watt/pub/reprints/2005-mc-charrec.pdf>. Accessed 20 Dec 2020
75. Xie X, Watt SM (2005) Recognition for large sets of handwritten mathematical symbols. Eighth international conference on document analysis and recognition (ICDAR'05), 1–4
76. Xinyan C, Hongli Y, Xin W (2013) Handwritten mathematical symbol recognition based on niche genetic algorithm. Third international conference on intelligent system design and engineering applications, ISDEA 2013, 803–806. <https://doi.org/10.1109/ISDEA.2012.191>
77. Xuejin Z, Xinyu L, Shengling Z, Baochang P, Tang YY (1997) On-line recognition handwritten mathematical symbols. In: Proceedings of the Fourth International Conference on Document Analysis and Recognition, vol. 2, pp 645–648. <https://doi.org/10.1109/ICDAR.1997.620585>
78. Zanibbi R, Blostein D (2012) Recognition and retrieval of mathematical expressions. Int J Doc Anal Recognit 15(4):331–357. <https://doi.org/10.1007/s10032-011-0174-4>
79. Zanibbi R, Blostein D, Cordy JR (2001) Baseline structure analysis of handwritten mathematics notation. Sixth international conference on document analysis and recognition, 768–773. <https://doi.org/10.1109/ICDAR.2001.953892>
80. Zanibbi R, Blostein D, Cordy JR (2002) Recognizing mathematical expressions using tree transformation. IEEE Trans Pattern Anal Mach Intell 24(11):1455–1467. <https://doi.org/10.1109/TPAMI.2002.1046157>
81. Zhang DY, Tian XD, Li XF (2010) An improved method for segmentation of touching symbols in printed mathematical expressions. IEEE International Conference on Advanced Computer Control, 2, 251–253. <https://doi.org/10.1109/ICACC.2010.5486679>
82. Zhao W, Gao L, Yan Z, Peng S, Du L (2021) Handwritten mathematical expression recognition with Bidirectionally trained transformer, *Handwritten Mathematical Expression Recognition with Bidirectionally Trained Transformer*
83. Zhu B, Nakagawa M (2011) On-line handwritten Japanese characters recognition using a MRF model with parameter optimization by CRF. Proceedings of the international conference on document analysis and recognition, ICDAR, may 2014, 603–607. <https://doi.org/10.1109/ICDAR.2011.127>