Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD38547

Artificial Intelligence Recognition of Children's Handwritten English Letters: Accuracy, Error Patterns, and Educational Applications

Yexin Zhang

Beijing Royal School, Beijing, China
zhangyexin2010@outlook.com

Abstract. Recognition of standard handwriting is efficient, but children's English letters are challenging because their shapes are under development. They have different directions, sizes, spacing, positioning, and orientations of strokes. All this makes the subtle differences between similar letters unnoticeable for recognition algorithms. This review covers the accuracy of recognition, frequent mistakes, and educational applications. Numerous research show impressive results on children's handwriting recognition in controlled conditions. However, recognition depends on case, regularity, quality of images, and the way of testing. The main problems are similar letters such as b-d, p-q, u-v, and i-j-I-l. Small mostly privately collected child datasets also reduce the possibility of comparison and application in classrooms. Recognizing algorithms could be used to detect attempts of practicing, provide immediate feedback on formation and measure the development of children. To use them safely, there must be confidence thresholds, teacher reviews, user-friendly interfaces and strict data protection rules. Now, these systems are good for controlled practice but not for assessment or diagnosis.

Keywords: children's handwriting, English letters, handwriting recognition, educational technology

1. Introduction

1.1. Background of artificial intelligence handwriting recognition in education

Artificial intelligence (AI) handwriting recognition converts handwritten content into machine-readable labels. The offline approach evaluates an image that is fully drawn. The online approach also takes account of the stroke sequence, timing, direction, and sometimes pressure. Such information provides insights into the way a letter was written by a child and not just what it looks like.

Handwriting recognition progressed from manually designed feature sets to learning approaches. Convolutional neural networks (CNNs) automatically learn visual features from examples. The EMNIST dataset, which is an extension of the Modified National Institute of Standards and Technology (MNIST) dataset, is useful for the evaluation of letters on a large scale, even though

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD38547

there are different class definitions among EMNIST variants [1]. Modern education-oriented systems utilize CNNs, tablet traces, augmentation, and mobility [2-4].

Education-oriented recognition should evaluate the prompted letter and give meaningful feedback. Available systems can recognize the missing letter, extra stroke, wrong shape, direction, sequence, and baseline [3, 5]. Thus, recognition assists rather than replaces the educational process.

1.2. Why children's handwritten English letters are difficult to recognize

Children's letter knowledge and motor control are still developing; hence repetition results may be different. Examples of mistakes include tilted stems, incomplete loops, lack of dots, misalignment, and inaccurate connections [6]. Such alterations affect both appearance and pen movement.

Most lower-case letters have common elements. Pairs of b-d and p-q are differentiated only through their orientation; u-v through their base curve; and i-j-I-l by dots, tails, or size. Letter reversals cause a discrepancy between visual recognition and the intended instruction.

Images in classroom settings are subjected to lines, shadows, erasure, other writings, contrast, and distortion from photography. Tablet-based input removes some of the noise but is different from pen-and-paper. Accuracy is thus affected by the individual, label set, medium, and context.

1.3. Research questions

The review addresses three research questions (RQs).

Research Question 1 (RQ1): How accurately can current artificial intelligence (AI) systems recognize children's handwritten English letters?

Research Question 2 (RQ2): Which English letters are most commonly misrecognized, and why?

Research Question 3 (RQ3): How could handwriting recognition be used to support children's English learning?

1.4. Scope and structure of the paper

This selective review addresses letters in the English and Latin alphabets used by pre-school and school-age children. Other scripts are included only where necessary to explain technology and data limitations. Sections 2-5 address methodology, accuracy, errors, and applications to education; Section 6 addresses the research questions.

2. Literature review

2.1. Method of selecting studies

The review includes twelve sources provided as well as additional searches carried out in the Education Resources Information Center (ERIC), arXiv, the National Institute of Standards and Technology (NIST), the Institute of Electrical and Electronics Engineers (IEEE), and journal publishers' databases. Most sources are from January 2017 to September 2026; older literature has been kept where relevant.

Search keywords included "children handwriting recognition," "English letters," "CNNs," "EMNIST," "letter reversals," "feedback," "tablets," "privacy," and other terms where appropriate. Papers had to either have an architecture and dataset, research on children Latin-letter writing, a learning tool for children, or a definition of the AI requirements for children.

DOI: 10.54254/2755-2721/2026.LD38547

Papers involving children's participants, with sample sizes, held-out data, and relevant performance metrics have been assigned more weight. General benchmarks offer context only. One Arabic source offers grounds for adult versus child comparison, but not for English letter accuracy. Due to differences in methodologies, the review applies a narrative synthesis approach.

2.2. Basic development of handwriting recognition technology

2.2.1. Traditional image processing methods

The traditional offline systems involve cleaning and normalizing the image, extracting features like contours and intersections, and classification using support vector machines and the like. These procedures work efficiently but rely heavily on the rules that do not work well when applied to the open loop, unusual stroke order, or misalignment of children's handwriting.

2.2.2. Deep learning methods

CNNs detect edges and shapes through examples and surpass many manually engineered systems. CNN accuracy goes from the baseline accuracy for the original EMNIST Letters data set of 85.15 percent to 93.44 percent in one study [1, 7]. One benchmark data set with large letters achieved 99.563 percent [8], but those data sets do not reflect the writing done by children.

Online CNN models include stroke sequences in the final output. Dynamic channels improved children's letter recognition [2]. The same type of information can be used to detect dysgraphia [9], but letter detection and diagnosis are different problems.

2.2.3. Recent artificial intelligence-based methods.

The latest systems use residual networks, augmentation, multi-channel input, confidence scores, and mobile inference. Educational systems could also compare a known target with acceptable examples. This constrained task supports formation feedback without inferring a child's intention from shape alone.

2.3. Previous studies on children's handwriting

Isolated online letters from 147 children have been investigated [2]. A system for teaching English using samples from children in four Asian countries was developed using a paper-and-tablet interface [4]. These experiments give direct experimental evidence for RQ1.

Quality assessment systems take into account more than just identity. Handwritten words of 231 young children have been examined [5]. Parameters like shape, direction, order, location, and movement have been studied using samples from 400 children [3]. Both the studies provide evidence for formation feedback.

Children's experiments have revealed difficult forms. The forms reported include b, d, i, j, l, p, q, u, m, n, w, x, and z [10-12]. Although not AI confusion matrices, these forms reveal the ambiguous forms a classroom AI must handle.

2.4. Common datasets and evaluation methods

EMNIST dataset is composed of 28×28-pixel images based on NIST. The ByClass variant of EMNIST consists of 62 classes of digits and letters while EMNIST Letters groups cases in 26

DOI: 10.54254/2755-2721/2026.LD38547

classes [1]. These variants cannot be compared because the task is simplified through merged labels.

Datasets for children are small and private. 27,000 online characters and 7,096 test characters have been used [2]. Multinational dataset which is not publicly available has been used [4]. This hinders replication and comparability.

Accuracy must be coupled with the letter-based precision, recall, F1 score (harmonic average of precision and recall), and confusion matrix. Word tasks must also consider character and word error rate. Evaluations for educational purposes must include false feedback, rejection of ambiguous inputs, teacher consensus, reaction time, and unassisted writing afterwards.

3. Recognition accuracy in previous studies

3.1. Reported accuracy in selected studies

Table 1 distinguishes child-specific findings from general benchmarking findings and cross-script findings.

Table 1. Selected handwriting-recognition results

Study	Population and input	Task and test set	Reported result	Interpretive limit
Cohen et al. [1]	General NIST writers; offline images	EMNIST Letters; 26 merged letter classes	85.15% baseline	Benchmark is not child-specific
Corbillé et al. [2]	147 children; online tablet traces	26 Latin letters; 7,096 test samples	95.00% best early-fusion CNN	Controlled tablet setting; private dataset
Zin et al. [4]	Kindergarten and primary students in four countries; paper and tablet	Uppercase, lowercase, mixed case, and 72-class variants	97.43% uppercase; 92.72% lowercase; 82.97% mixed case; 85.74% all classes	Accuracy changes with class set and normalization
Zin et al. [4]	Children's handwritten words	1,000 words; 4,607 characters	95.60% words; 98.76% characters	Segmentation and expected vocabulary affect results
Liu [7]	General EMNIST writers; offline images	26-class EMNIST Letters	93.44%	Not child-specific; limited model reporting
Saqib et al. [8]	General Kaggle A-Z images	26 uppercase classes; 74,490 test images	99.563%	Curated benchmark; does not estimate young children
Alhaeraki et al. [13]	Child and adult Arabic datasets; offline images	Same CNN tested across separate datasets	91% child dataset; 97% adult dataset	Different script; indirect evidence only

Note. The findings cannot be directly compared since the writers, labeling, inputs, preprocessing, and testing are all different.

DOI: 10.54254/2755-2721/2026.LD38547

The controlled child-specific studies are at about 95 percent accuracy, but the results vary according to the task. Best early fusion CNN performed at 95.00 percent accuracy in dynamic channels [2]. Uppercase letters were recognized at 97.43 percent, while lowercase was at 92.72 percent, and mixed cases at 82.97 percent [4].

Average accuracy can hide harmful errors. Even at 95 percent, one letter in twenty will be a mistake, and the mistakes cluster depending on letter, writer, or capture mode. Classroom tools therefore need calibrated confidence and must withhold feedback on uncertain inputs.

3.2. Differences between adult and children's handwriting recognition

General standards outperform children's criteria mainly because their pictures are more clean and consistent [2, 8]. It is incorrect to take high scores as evidence that the classroom performance of these students is better.

3.3. Factors affecting recognition accuracy

3.3.1. Uppercase and lowercase letters

The uppercase letters were easy to recognize compared to lowercase letters; while the mixed case were hard to recognize [4]. The lowercase letters are based on minor differences, whereas the mixed case have similar pairs like C-c, O-o, P-p, I-l, and S-s. Hence, the EMNIST dataset combines case labels [1].

3.3.2. Neat and irregular writing

Irregular writing includes slants, heights, loops, and strokes that vary from standard forms. However, augmentation works only if synthetic variations mimic the mistakes made by children. Models should be able to differentiate between recognized letters and those needing correction.

3.3.3. Clear and noisy image backgrounds

The crop and binarization processes can eliminate any dots, tails, or faint strokes. The ruled paper, shadows, and angles at which the image is taken introduce patterns that are not present in standard datasets. Classroom tests should therefore include ordinary devices, paper lines, erasures, and varied lighting.

3.4. Limitations of comparing results from different studies

Variations occur across studies due to unit differences, number of classes used, case strategies, input, writer split, and pre-processing methods. It is also possible that random image splits would put a writer both in the training and test set. Additional restrictions in terms of private data and single best-run results make comparisons even more difficult.

Classroom adoption should be based on the risk involved. While commendations can survive some level of abstaining, criticisms, assessments, and screenings need more proof. Results should be presented per letter, learner group, device used, and writing condition.

DOI: 10.54254/2755-2721/2026.LD38547

4. Common recognition errors and possible causes

4.1. Confusion between letters with similar shapes

None of the child-specific studies have universally established the order of difficulty of English letter errors. However, the repeatedly hard letter sets may be determined on the basis of classifier outcomes and children's letter reversals' research.

4.1.1. b and d

Lowercase letters b and d contain the same stem-and-bowl elements which are reversed. Letter reversals happen very frequently in children [10, 11]. Any learning system ought to warn about the possibility of a letter reversal and utilize stroke direction where it is known.

4.1.2. p and q

The p-q pair includes reflection and descender placement. Crop marks or an unfinished bowl contribute to ambiguity [10, 14]. Directionality may help determine how the letter was formed, but it will not fix any inversion problems in the end form.

4.1.3. u and v

A difference between u and v may arise from a minor bottom curvature, and u may also be mistaken for n. Research on child development has also shown that m, w, n, and u are difficult symbols [10, 12]. This should be incorporated in the training data.

4.1.4. The i-j-I-l group

i-j-I-l relies heavily on dots, tails, and vertical extension. Cropping and low contrast will lose these features. This has been proved by studies in children and EMNIST data sets [7, 10]. Preprocessing needs to retain small features.

4.2. Irregular letter size shape and spacing

Instruction is also affected by size and location. For example, b could become o with a short ascender and p could look like P with a short descender. Spacing could either separate letters that were meant to be one or could blend letters meant to be separate.

4.3. Limited datasets of children's handwriting

The publicly available datasets consist of either adult or combined-age writing, but the significant children datasets are not public [2, 4]. Scientists must use artificial distortions or small-scale local samples, which may not reflect the variation that is found in classrooms.

Dataset limitations may mask the differences in performance based on the age of the writer, his/her motor development, style of writing learned, language background, disability, used device, or the hand used for writing.

DOI: 10.54254/2755-2721/2026.LD38547

4.4. Differences between adult and children's handwriting

Adult handwriting is characterized by more consistent letter forms and motor control than that of children who are still developing knowledge about shape, sound, orientation, and movement. Their variability should be considered developmentally relevant, hence the requirement for realistic models.

5. Educational applications and limitations

5.1. Possible uses in English letter learning

5.1.1. Automatic letter identification

The model may examine if the attempt aligns with the letter presented in the prompt. In early activities, only the learned letters should be considered, and the child should be asked to confirm his/her intention if he/she lacks confidence.

5.1.2. Immediate learning feedback

The feedback must relate to a single observable aspect, like the position of the stem or dot. Previous work focuses on analyzing incomplete strokes, letter form, stroke direction, stroke order, and baseline alignment [3, 5]. Low-confidence attempts should receive no automatic criticism.

There have been claims of benefits of digital handwriting tasks based on small classroom experiments [15], but they cannot be considered conclusive evidence.

5.1.3. Practice and progress tracking

Progressive technology can record repetition and discover consistent difficulties with letters or baselines. Original writing must be kept next to the score so that teachers can differentiate between learning and changes in technology or identification.

5.2. Requirements for classroom use

5.2.1. Reasonable recognition accuracy

No single accuracy threshold ensures classroom safety. Validation must be done through independent writers, letter-by-letter accuracy, realistic devices, confidence calibration, and false-feedback rates. The system must refrain from acting on ambiguous inputs, and teachers must evaluate high-stakes actions.

5.2.2. Simple operation

A child interface needs to have minimal steps and be capable of accepting typical school writing. Mobile offline recognition technology may help eliminate delays and unnecessary transmission of data [4] even though stored data will need security measures.

DOI: 10.54254/2755-2721/2026.LD38547

5.2.3. Protection of children's data and privacy

Handwritten information may disclose the identity of the writer and his/her learning difficulties. As per guidelines from the United Nations Children's Fund (UNICEF), issues such as privacy, safety, equity, transparency, and children's best interest must be taken into account [16].

5.3. Current limitations

Data is limited because of privately owned data, small studies, labeling inconsistency, and controlled experiments. Some models might accept misformed letters or reject correctly formed ones. Results of tablets cannot be easily applicable to photographs on paper, and context may obscure misformed letters.

Bad or repeated feedback might lower the level of confidence. Feedback lacking an explanation does not help in formation, and recognition mistakes should not be regarded as diagnosis [17].

5.4. Suggestions for future educational tools

The future tool should first evaluate the quality of input, then determine the intended letter, compare it against the lesson target, provide only one suggestion for correction, and record only the progress information needed. The teacher should be able to review each decision made.

Tests should include whole learning systems, involving new writers, realistic graphics, multiple devices, and regionally acceptable letters. Both immediate performance and later independent writing on paper should be measured, and evaluation guidelines published, even if child performance data cannot yet be released.

Adaptive practice may involve repeating difficult letters, but not labeling children nor diagnosing disorders. Teachers need to regulate regionally acceptable styles, override the feedback, and review the basis for suggestions.

6. Conclusion

6.1. Answers to the research questions

RQ1: Current systems work well for controlled child-specific tasks, yet performance varies based on case, labels, authors, devices, and image quality. Performance in natural classroom conditions is unknown. Studies using tablets in controlled settings still cannot be seen as evidence of reliable performance in normal paper environments or uncontrolled classroom handwriting.

RQ2: There is no general ranking of errors. Groups of high risks are: b-d, p-q, u-v and other rounded letters, i-j-I-l. Reversals, underdeveloped shapes, differences in size and preprocessing will erase the unique properties of those errors. Since child-specific confusion matrices are not often published, ranking should be performed for each dataset and each letter separately.

RQ3: Recognition is helpful for detecting attempts, providing targeted feedback, giving guidance for practice, and monitoring progress. It is supposed to be used in a supervised way, as a tool for withholding judgement. Autonomous grading or diagnosis require much stronger validation. Teachers should keep control over feedback and grading.

DOI: 10.54254/2755-2721/2026.LD38547

6.2. Main findings and contribution

Accuracy, child recognition, and educational effectiveness are distinguished in the review. Evidence is strong for standardized data sets, moderate for writing by children, and tentative for educational impact. The effect of class definition may be as important as that of the model design. The principal contribution consists of the three-fold synthesis of the issue that explains why good benchmark scores cannot be easily transposed into the classroom setting.

6.3. Limitations of the paper

This is not a systematic review but a narrative review. Two of the selected studies explicitly provide accuracy results for letters of the English or Latin script written by children. Other evidence is from related tasks or writing systems. It is impossible to combine the data due to heterogeneity, unavailability of data, and lack of confusion matrices. The source selection requires subjective judgment. There is no new data collected in the paper.

6.4. Possible directions for future research

A common method needs to use age-labeled, writer-independent samples; capital and small letters; online traces; and lifelike pictures on paper. The results need to show per-letter statistics, confidence, rejected stimuli, agreement between teachers, and group equity. In education, trials are required to assess later unaided handwriting and agreement between teachers.

References

[1] Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks* (pp. 2921-2926). IEEE. https://doi.org/10.1109/IJCNN.2017.7966217

[2] Corbillé, S., Fromont, É., Anquetil, É., & Nerdeux, P. (2020). Integrating writing dynamics in CNN for online children handwriting recognition. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 270–275). IEEE. https://doi.org/10.1109/ICFHR2020.2020.00057

[3] Hamdi, Y., Akouaydi, H., Boubaker, H., & Alimi, A. M. (2022). Handwriting quality analysis using online-offline models. *Multimedia Tools and Applications, 81*(30), 43411-43439. https://doi.org/10.1007/s11042-022-13228-w

[4] Zin, T. T., Thant, S., Pwint, M. Z., & Ogino, T. (2021). Handwritten character recognition on Android for basic education using convolutional neural network. *Electronics, 10*(8), Article 904. https://doi.org/10.3390/electronics10080904

[5] Simonnet, D., Girard, N., Anquetil, E., Renault, M., & Thomas, S. (2019). Evaluation of children cursive handwritten words for e-education. *Pattern Recognition Letters, 121*, 133-139. https://doi.org/10.1016/j.patrec.2018.07.021

[6] Doğan, M. F., & Doğan, Z. (2018). Identifying common errors in vertical lowercase manuscript writing of the first graders in primary school. *Journal of Education and e-Learning Research, 5*(3), 144-156. https://doi.org/10.20448/journal.509.2018.53.144.156

[7] Liu, H. (2023). Handwritten English character recognition using convolutional neural network. In *Proceedings of the 3rd International Conference on Signal Processing and Machine Learning* (pp. 199-204). https://doi.org/10.54254/2755-2721/4/20230450

[8] Saqib, N., Haque, K. F., Yanambaka, V. P., & Abdelgawad, A. (2022). Convolutional-neural-network-based handwritten character recognition: An approach with massive multisource data. *Algorithms, 15*(4), Article 129. https://doi.org/10.3390/a15040129

[9] Zolna, K., Asselborn, T., Jolly, C., Casteran, L., Nguyen-Morel, M.-A., Johal, W., & Dillenbourg, P. (2019). The dynamics of handwriting improves the automated diagnosis of dysgraphia. arXiv. https://doi.org/10.48550/arXiv.1906.07576

DOI: 10.54254/2755-2721/2026.LD38547

[10] Richmond, J. E., & Taylor, M. (2014). Visual recognition difficulties: Identifying primary school learners' directional confusion in writing letters and numbers. *South African Journal of Occupational Therapy, 44*(3), 2-6. Retrieved from https://journals.assaf.org.za/index.php/sajot/article/view/19668

[11] Fischer, J.-P., & Luxembourger, C. (2021). A test of three models of character reversal in typically developing children's writing. *Frontiers in Communication, 6*, Article 719652. https://doi.org/10.3389/fcomm.2021.719652

[12] Adhe, K. R., Mustaji, Suprapto, N., Suryanti, & Ling, L. Y. (2024). Difficulty of visual recognition: Identifying the direction confusion of reading letters in young children. *International Journal of Education in Mathematics, Science and Technology, 12*(2), 334-344. https://doi.org/10.46328/ijemst.3888

[13] Alheraki, M., Al-Matham, R., & Al-Khalifa, H. (2022). Handwritten Arabic character recognition for children writing using convolutional neural network and stroke identification. *arXiv*. https://doi.org/10.48550/arXiv.2211.02119

[14] Subramaniam, V., & Nasir, N. S. A. (2020). Multisensory therapy in letter reversal of dyslexic pupils. *Universal Journal of Educational Research, 8*(12), 7118-7130. https://doi.org/10.13189/ujer.2020.081279

[15] Tang, J. T., Chu, S.-T., & Chang, T.-F. (2024). Enhancing English alphabet handwriting skills in preschool children through digital game-based learning approach. *Innovation in Language Learning and Teaching*, 1-19. https://doi.org/10.1080/17501229.2024.2352791

[16] UNICEF. (2025). Guidance on AI and children: Version 3.0. *UNICEF Office of Strategy and Evidence Innocenti*. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

[17] Liu, S., Rahim, N., & Ying, S. (2026). Process-based digital assessment of children's handwriting: A scoping review of graphomotor, orthographic, and educational indicators. *Frontiers in Education, 11*, Article 1923428. https://doi.org/10.3389/feduc.2026.1923428