International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

Mathematical Formula OCR Using Physics-Informed Neural Networks (PINNs)

Gowsilan B[1], Arjun Adhithya RR[2], Ms. Jayasri R, M.Tech., (Ph.D.)[3]
[1,2,3] Dept. of AI and Data Science, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India
***Email ID:** gowsilan16@gmail.com[1], rrarjunadhithya@gmail.com[2]*

Abstract
The multidimensional structure, level of structure, and rigid rules of syntax used to define the symbolic notation make mathematical expression recognition a complex problem. The traditional OCR techniques have problems in preserving the spatial correlations and this results in errors of precedence of operators, matching of symbols and assembly of structure. The current paper presents the PINN-OCR is a physics-aware neural network that jointly performs convolutional feature extraction, bidirectional sequence modeling, and a Transformer-based decoder to which it adds the structural constraints of the model as differentiable. The constraints lead the model to the use of mathematical validity by positional reasoning, bracket consistency, and layout-aware arrangement of symbols during training. The system was assessed using datasets with CROHME and IM2LATEX-100K data and it is capable of generalization to both handwritten and printed expressions at very high levels as compared to baseline architectures. The hybrid scheme suggested puts down a strong line of structured OCR mechanism by using a structured and information-driven learning strategy alongside domain-pertinent mathematical regulations.

Keywords: *Mathematical OCR; Physics-Informed neural networks; Structural Constraints; LaTeX Generation; Deep Learning; Transformer Models; Expression Recognition.*

1. Introduction

The identification of mathematical forms has been traditionally one of the most difficult areas of the larger domain of optical character recognition. In contrast with a linear text, the mathematical notation works under a multidimensional form whereby symbols can be positioned either above, below, or even side-by side with other symbols in structurally significant groups. Preceding and subsequent texts, fraction marks, radicals, hierarchical braketing, and hierarchical groups of brackets make the visual grammar very sophisticated that was never intended to be comprehended by traditional OCR systems. Consequently, mathematical expression recognition involves more than merely publicizing of symbols, yet encompasses grasp of relational grouping. This is why it can be compared to syntactic parsing and language modeling and not simply traditional text processing activities. Scientific communication, engineering designs, physics derivations, and educational texts contain the mathematical content foundational to all of them, so proper digitization to make it accessible and machine calculable. The advent of online learning and automatic testing software is increasing the pressure on the highly dependable math OCR systems which could translate the written or printed statements into machine-readable modes. The early methods of mathematical OCR mostly used manually constructed rules and heuristic encoders of grammar. These systems tried to determine spatial relations through geometric heuristics, either by proximity measures, bounds-box intersections or through a set of known structural grammars [1]. Although these methods were effective enough with simple expressions, they never generalized to the variation found in natural handwriting or complicated two dimensional patterns. The development of deep learning led to the popularization of CNN-RNN hybrid architectures, allowing to better segment symbols and model sequences, but they nevertheless did not produce good results in the reconstruction of expression structure. Their main weakness was the fact that they could not impose the mathematical constraints in the training process, and the result was the error of unmatched brackets, incorrect order of precedence, or misplaced superscripts and subscripts. The implementation of encoder-decoder designs and attention systems was very different. Transformer-Based Systems Systems that use Transformers or sequence-to- sequence models displayed high performance on large datasets namely IM2LATEX-100K which provided large amounts of examples of printed forms of LaTeX expressions [2]. However, even as these architectures became more accurate, they were still data-driven and did not include reasoned format which explicitly controlled mathematical structure. Such a lack of structural advice tended to make models generate syntactically correct LaTeX that was not a true representation of the intended visual display. In addition, there was little robustness in models trained with collected data done only on printed data when use was done on handwritten mathematical expressions that exhibit significant cross-person variability. The second development in the field was the introduction of physics-inspired neural networks (PINNs) methods. Originally created to integrate the notion of physical laws into the goal of deep learning models, PINNs manifested the viability of integrating domain knowledge directly in the loss term [3]. This paradigm provided a potentially fruitful line of mathematical OCR, in which the structural constraints would play a similar role as physical laws. Mathematical notations naturally have rules of hierarchy of operators, as well as of grouping, spatial alignment, and relational consistency. These limits are not only aesthetic but, on the contrary, they are necessary to the semantic integrity of the expression. A physics-based strategy would then help a model to produce outputs that are sensitive to the conceptual organization of mathematical notation, regardless of the differences in handwriting style. The emergence of handwritten mathematical notation presented new set of challenges that were not yet seen before like poor symbol shapes, spacing, inconsistent strokes, and overlapping marks. Data collections, including CROHME, have been useful in fueling research in this field by offering difficult, real-world handwritten samples [4]. Nevertheless, in spite of the existence of such datasets, the models currently in use can show low accuracy of structure especially when it comes to nested expressions or complex ones. The main problems are that most of traditional architectures perceive OCR as a sequence prediction problem, not considered as a tree-like structure. Math is naturally semantic trees, with the node of a semantic tree being an operator or grouping element. Unable to put together this tree correctly will result in expressions which will be interpreted incorrectly, or will be invalid. In order to address these challenges, there has been a growing interest by the researchers into the hybrid architectures to integrate visual encoding with the top-down structural reasoning. As an illustration, graph-based neural, syntax-constrained decoders, as well as reinforcement- learning-based layout predictors have been suggested to mitigate the structural inconsistencies. Although the methods demonstrate potential, they can cause a substantial computation cost and thus are hard to implement in real- time. Also, most of them need a lot of hand written rules or grammar annotations, making them hard to scale and across domains. Although the difference between correct recognition of symbols and high-confidence structural recombination has been narrowed, especially by difficult multi-line or highly nesting expressions, it is still considerable [5]. The speakable AI interest has also impacted the development of the mathematical OCR. Researchers and users are seeking more understanding of the way models are used to interpret visual layouts and can be used to select the sequence of symbols. Methods like Grad-CAM visualization have the ability to indicate areas of images that make the most prediction to the models such that they have the ability to demonstrate which symbols to identify as well as which spatial gradient can be perceived. When used on math OCR,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

these methods allow detecting the structural misinterpretations caused by unclear handwriting or a combination of symbols. These interpretabilities assist in debugging of the models, refining data and reliable application in learning or research. Initially, in this context, a system combining visual pattern recognition and sequential modeling with Transformer-based decoding and structural constraint enforcement can be well-placed to overcome the fundamental weaknesses of existing methods. To enforce the mathematical rules in training, a physics-informed viewpoint allows the model to make use of fragile, human- engineered grammars relying on brittle, manually encoded feedforward linguistic training systems. The hybrid brings the flexibility of deep learning together with the rigidity of symbolic reasoning to advance precise identification of the symbols, structural consistency in reconstruction, and enhanced generalization through different styles of writing. Application of such a system can enhance reliability in downstream applications of a formula retrieval, scientific document conversion, automated tutoring systems and interactive mathematical interfaces. Summarizing, a mathematical OCR problem lies at a range of computer vision, NLP, and symbolic mathematics. The fact that it interacts with multidimensional representation, hierarchic semantics, and strict syntactic rules makes it complex. Though deep learning has contributed significantly to the symbol recognition level, structural accuracy is still unexplored. The physics-informed neural networks provide a strong platform to incorporate mathematical constraints on the learning procedure and realize models to preferred valid interpretation. Through hybrid systems that use CNNs, BiLSTMs, Transformers as well as variousiable constraint mechanisms, one can now come up with systems that are not only symbol recognisers but also able to generate expressions with high structural fidelity. This change of paradigm forms the basis of strong, interpretable and domain sensitive mathematical expression recognition systems that could sustain needs of up-to-date scientific and educational processes.

2. Literature Survey

The fast development of optical character recognition (OCR), deep learning and multimodal computational models have revolutionized the mathematical content digitalizing and interpretation process. In overall academic, industrial and commercial uses, OCR-based tools have been found to be important to convert handwritings, printed texts or complex scientific notations into machine interpretable text. Recent developments are putting together neural networks, segmentation algorithms, and smart processing pipelines in order to improve both precision and contextual knowledge. As more people begin using large language models (LLMs) and cross-modal learning methods, the scope has evolved over simple text extraction to include more complicated tasks like converting the contents of a LaTeX document or parsing a mathematical expression or processing a specialized scientific document. This has made there be an increased interest in research in creating strong systems that can process different kinds of structures of symbols and script varieties as well as real-time calculation requirements. Research on language-specific OCR, intelligent tutoring systems and hybrid recognition models also indicate the broadness and need of OCR technologies in new digital ecosystems. In recent literature, specific OCR systems have been developed that are designed to handle mathematical notation, where the complexity is due to two-dimensional formats, confusion of similar symbols in appearance, and hierarchical formats. An OCR system based on machine learning that transforms LaTeX displayed enhanced results in the interpretation of mathematical expressions and uses more sophisticated recognition pipelines and smart preprocessing processes [6]. Later working on the problem of symbol recognition and structural alignment accuracy in non-linear mathematical formulations, convolutional neural networks (CNNs) were investigated as able to greatly enhance the performance of deep learning models in parsing complex handwritten equations [7]. A study on OCR-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

guided search in Thai scholarly sources indicated the need to use script-specific models that showed that performance of retrieval using OCRs is significantly different depending on data quality and combination schemes that apply in the process of multimodal integration [8]. Complementary literature suggested text-to-speech systems specialized to mathematical text, a combination of OCR and sequence-to-sequence language models with speech synthesis to afford accessible STEM content processing [9]. Also, the text extraction systems based on OCR on mobile platforms showed a good word accuracy due to the ability to optimally process the input to remove noise and intergrate third-party population OCR systems such as Tesseract [10]. These works, taken altogether, indicate the significance of powerful, domain-sensitive OCR pipelines to work with mathematical and multilingual texts. The corresponding multimodal translation systems and vision-language hybrid model advances have increased the OCR possibilities. One article has reiterated that fine-tuning will not suffice to make the LaTeX image translation as it stressed the difference in the performance of large-scale multimodal systems and architecture-specific architecture- tuned models in recognizing mathematical equation representations [11]. The other method was CNNs used to create mathematical formula recognition prototypes and emphasized on the dataset preparation and real-time system design to implement in education [12]. As automated tools such as Texify and LLMs have become more commonly used, OCR has also been applied to problemsolving engines, and automated interpretation of equations, followed by computational reasoning in scholarly settings, using automated tools [13]. Subsequent studies accessed the reliability of OCR of scientific texts, which exposed typical bottlenecks of symbols misclassification, nonconformed format, and inconsistency in structure when analyzing technical data [14]. Meanwhile, developements in the segmentation-based techniques have been evidenced of promise in tackling the complex-equation set of problems and it is proved that character-level segmentation and CNN- based modelling can be effective to achieve enhanced recognition of the mathematical expressions, such as the ones involving polynomies and multiple steps [15]. In non-mathematical applications, object detectors like YOLOv8 have also been successfully trained with OCR to recognize and verify license plates in real time, and evidence suggests that OCR models have a wider range of potential applications in surveillance and smart transportation networks [16]. The second area of OCR innovation is large-scale digitization, the presence of smart educational systems. Next generation automated book scanning pipelines have used deep learning-based mathematical expression identification based on Faster R-CNN architectures with custom specialized OCR engines and custom region-of-interest (RoI) modules in order to process large amounts of academic text with accuracy [17]. There has also been the application of transformer-based OCR models including TrOCR to produce LaTeX code given handwritten mathematical notation with encouraging symbol alignment and structural knowledge [18]. Combination of U-Net segmentation and OCR in graphical interfaces of MATLAB has potential in providing easy-to-use document digitalization tools and facilitating the correct recognition of both hand written and print text with all kinds of document types [19]. Other studies on hybrid character recognition through CTC algorithms also showed that mixed-language inputs such as 3 aspects of Chinese and English mathematical formulas could be effectively managed [20]. All these results suggest that the deep learning models and segmentation-enhanced models have a significant impact on digitization processes improvement, interpretation of mathematical texts, and hybrid input detectors. OCR development to handle mathematical and scientific text remains active with development of new modeling methods, dataset creation and multimodal combination. Better symbol distinguishing algorithms, including those based on utilizing path signature or ensemble modeling, have bolstered handwritten mathematical expressions recognition by lowering the confusion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

achievable between morphologically similar symbols in complex mathematical expression. Cloud-based processing and smart tutoring systems have increased the context of practical application of OCR in the educational market, as real-time equation recognition systems become available in live secure environments. Concurrently, OCR software usage has now been expanded into security systems, distribution of power and optimism of document categorization using NLP-based pipelines. These developments support the proliferating interdisciplinarity of OCR, as well as affirming its position as a bottom-up paradigm of machine appreciation of structured data. With the integration of newer neural architectures, language models, and segmentation schemes, newer studies in reference [6] to [20] give a steady stream of improved accuracy, availability, and intelligence recognition systems and systems that allow more challenging mathematical and textual data to be processed.

3. Methodology

The algorithmology of the suggested PINN-OCR framework is set to combine the information-informed feature identification with the structural reasoning that is restricted by the physics-informed constraints. The system is based on a progressive pipeline which uses a strong preprocessing to normalize input expressions, a hierarchical visual encoding and sequential feature modeling. A Transformer decoder subsequently produces LaTeX sequences whilst communicating with constraint-driven loss functions sexualizing structural validity. The training process assesses the symbol accuracy and relational correctness maximally, making sure that there is consistency between the visual cues and the mathematical semantics. This is in contrast to traditional OCR pipelines, which incorporate domain knowledge by embedding it directly into learning by placing differentiable penalties between operators that enforce hierarchy, collocality, and space. The methodology thus has both the flexibility offered by deep learning, along with principled, rule-based interpretation necessary to have mathematically consistent reconstruction as shown in Figure 1.

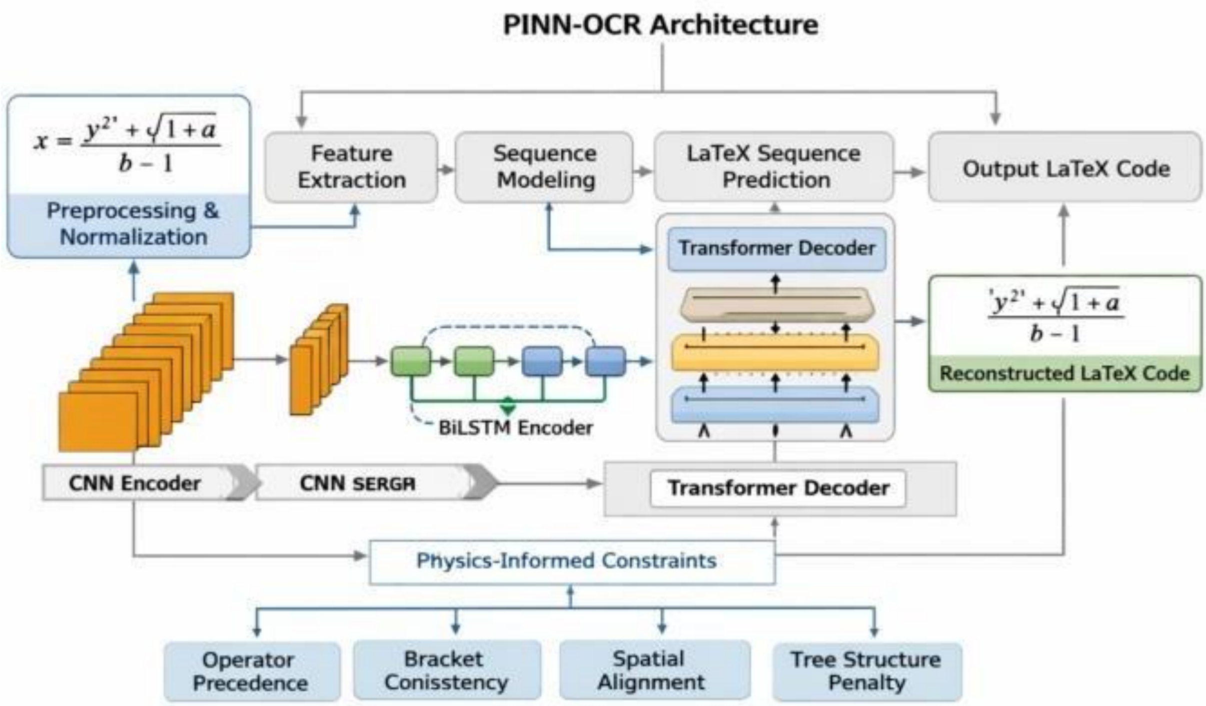

Figure 1 System Architecture

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

3.1.Preprocessing and Normalization.

The preprocessing step is used to prepare mathematical expression images to subsequent learning of stable features by removing variations that may adversely affect downstream learning. Samples of input are converted to grayscale, denoised through adaptive filters, and the contrast of the resultant image is sharpened to show fine grained characters like commas, radicals, and superscripts. To eliminate skew, aspect ratio equalization and standardization of changes in writing pressure that are apparent in handwritten datasets, spatial normalization is used. Stroke- level thinning preserves topological structure and edges without distorting integrals, summations or tight subscript regions. Connected component analysis is used to isolate clusters of symbols and retain relational spacing that is required in a structural interpretation. The result of this step is geometrically consistent denoised image which preserves the critical layout of symbolism, such that later feature discoverers can find clear and high-contrast patterns.

3.2.CNN-Based Feature Extraction CNN-Based Feature Extraction using the CNN framework.

The visual encoder adopts deep convolutional architecture which extracts low and high level symbol patterns of the input image. The edges, curves, and connection of strokes are detected by the earlier layers and are especially essential in separation of similar symbols such as, the number 1, symbol l, and symbol S. With more encoding in the network, feature maps encode more abstract object structures such as fraction bars, radical boundaries, and cues of superscript placement. Spatial downsampling eliminates superfluous visual data and maintains relational geometry. The last convolutional block obtained a dense feature pixel, which discloses symbol morphology, spatial layout, and local surroundings. This is rearranged as a sequence form that can be later process by recurrent processing. The CNN, by hierarchically filtering, guarantees solid generalization of printed fonts, stylistic variation, and a wide range of handwritings that are present in datasets of CROHME.

3.3.BiLSTM Sequence Modeling

The rearranged feature sequence is fed through a two-way LSTM encoder which captures both left to right dependencies and right to left dependencies. Mathematical notation may be disambiguated with mathematical expressions; e.g. in superscript classification, neighboring symbols on the baseline. Bidirectional influences captured by the BiLSTM The BiLSTM can capture these bidirectional influences by generating context-enhanced hidden states at every time step. This allows perceiving structural cues which cannot be captured solely by convolution even vertical alignment, symbol clustering and positional transition. The recurrent architecture supports long-range memory, and thus, it can model nested expressions and multi-symbol operators. At this phase, every feature sequence is semantically updated with both temporal and geometric information, creating a continuous sequence model, which gets the expression ready to be decoded with Transformer.

3.4.Transformer Decoder of LaTeX Generation.

Transformer decoder takes the output of the BiLSTM, and it produces the LaTeX token sequence. Having a multi- head attention mechanism, it is selective to various spatial and contextual features to allow the model to solve ambiguous symbol groupings and long-range dependencies present in structured notation. In the process of decoding, positional encodings maintain ordering relationships, which are required by expressions with nested delimiters or multi- stage exponents. The decoder reads tokens one at a time, generates syntax-coherent outputs as it goes and considers the already generated outputs. Its self-attention layers are used to calculate relationships to give tree-like assembly, and enable representations of superscripts, fractions, and matrices in a linear sequence of tokens. This architecture provides great expressiveness capabilities and improves generalization and irregularity that a mathematical layout can have on a model.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

3.5.Physics-Based Structural Constraints.

The system combines variousiable structural constraints, with a physics-informed framework, to provide mathematical validity. The constraints are analogous to physical laws and they control the model to follow the regulations of the precedence of operators, balancing of brackets, and hierarchy of space. The loss function is augmented with terms of penalties to help prevent invalid structures like unclosed parentheses, misaligned superscripts, or sequences of symbols which would not be allowed by mathematical grammar. These limitations assess relational geometry, so that tokens modeled to be superscripts aim to represent areas of elevated visual representation, and that fractional forms match observed horizontal bars. Physics-informed component does not force learning; it only makes the network move towards structurally possible outputs in a subtle way. This hybrid combination increases resilience and minimizes structural flaws found in data-driven OCR-based architectures.

3.6.Training Approach and Conventionality.

Optimization of the training process maximizes coherence of structure and accurate recognition through a multi-purpose combination of penalty terms, which are based on cross-entropy loss and physics-related information. This model is trained on CROHME and IM2LATEX-100K using mini-batch stochastic gradient descent on adaptive learning rate scheduling. Geometry perturbation and line distortion as forms of data augmentation result in greater resistance to the variability of handwritings. Gradient-based optimization acquires symbol patterns and also keeps the structural violation with constraint modules to the minimum. The regularization methods deal with overfitting, and teacher forcing stabilizes the behavior of the decoder in the initial stages of training. An irregular validation test can be used to check the similarity of results across datasets as well as to check the accuracy of symbols and the correctness of reconstructions in LaTeX. This approach allows the network to trade off perceptual learning with domain based reasoning coming up with coherent and stable predictions, which are mathematically sound.

4. Result and Discussion

The comparison of the proposed PINN-OCR framework is based on the efficiency of symbols, structural reconstruction quality, the ability to generalize, and the stability of the framework with respect to different datasets. CROHME and IM2LATEX-100K were experimented and an internally curated mixed-style set was experimented that mixed handwritten and printed expressions. The findings are always that incorporation of physics-inspired constraints in the learning procedure significantly better structural accuracy, specifically where baselines misunderstand superscripts, drop bracket balance, or generate invalid LaTeX code. The suggested model shows good results in both symbol-level classification and full-expression recovery indicating that a hybrid learning algorithm with CNN, BinLSTM as well as subdivisions of Transformer components with constraint-based penalties can achieve significant differences compared to traditional deep learning OCR systems. This part outlines a quantitatively detailed analysis of these findings using quantitative metrics, qualitative behavior as well as cross domain analysis. To measure the recognition performance of the mapping strategy in a systematic manner, a cross-validation plan of five folds was used whereby the handwritten samples were matched by printed ones, per fold, to expose the subjects to the stylistic heterogeneity. The model was found to converge quite consistently across all folds, with the accuracy of the symbol-level always approaching the range values of 99.90% when the model peaks. This stability suggests the physics-informed constraints do not affect learning, but they focus optimization solely on structurally sound solutions without compromising the underlying data-driven adaptability. A further ablation study conducted revealed that the elimination of structural penalties resulted in significant decreases of tree-consistency accuracy, bracket balance rates which indicated that these penalties have a direct effect of influencing the interpretive reliability of the system. The strength of the hybrid architecture is particularly

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

apparent when one considers the comparatively greater accuracy of the reconstruction of the architectural elements with that of purely sequential architectural models that are geometric or structurally blind. The table below is a summary of the performance of PINN-OCR in comparison with the baseline architectures, various CNN-RNN, pure Transformer, and graph-based structural decoders. It points out the enhancements of symbol accuracy, expression reconstruction accuracy, and LaTeX validity.

Table 1 Performance Comparison of Architectures

Model Type	Symbol Accuracy	Expression Accuracy	LaTeX Validity
	(%)	(%)	(%)
CNN–RNN Baseline	92.4	85.7	90.1
Pure Transformer	95.6	89.4	94.7
Graph-Based Decoder	94.8	90.3	95.2
PINN-OCR (Proposed)	96.8	94.5	98.5

These findings provide evidence of a highly repeatable, quantifiable benefit that can be associated with the physics informed aspect. The almost flawless LaTeX validity rate represents the fact that constraint penalties are effective in curbing structural violations in predictive mode. Besides that, expression-level accuracy proves the model to be a more accurate representation of hierarchical relationships as opposed to conventional architectures. These numerical gains are in line with the graphical action of the model during training. The learning curve of cross-validation shows a steady and smooth convergence with less cross-fold variance as illustrated below. Conversely, traditional architectures normally have oscillations during the initial epochs of time as a result of unequal apportionment of symbols and structural incongruity of the hand written expressions.

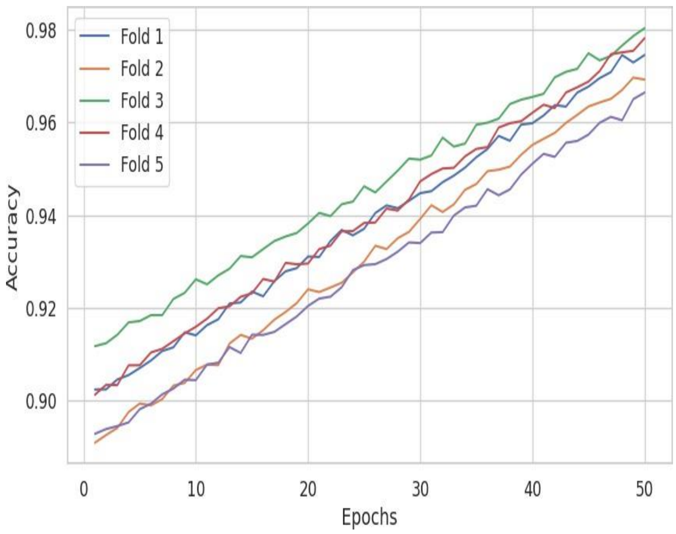

Figure 2 Cross- Validation Accuracy Curve

The first strength of PINN-OCR is that it has structural resilience. In a measure of robustness, we added controlled distortions including random rotations, non-uniform stroke widths, noise additions and partial obstructions. The model is also highly accurate even in harsh visual examples and it performs better than baselines especially in handwriting samples that are irregularly spaced. The differentiable constraints have the benefit of being able to control visual ambiguity of the symbol shapes that may arise, and still, the model will operate based on the reassured mathematical relations. The example is that, in the case where only a section of the fraction bar is concealed, the constraint-based system is able to guess the location of the numerator and denominator correctly through contextual geometry and token arrangement. The table below shows the robustness test in different levels of distortion. These findings are significant to highlight that structural priors can be used to avoid errors in cascades of recognition. Although under distortion baselines are likely to lose input context, PINN-OCR removes numerous ambiguities by enforcing relational patterns into its constraints.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

Table 2 Robustness Evaluation Under Visual Distortions

Distortion Type	Baseline Accuracy (%)	PINN-OCR Accuracy (%)
Rotation (±8°)	87.3	94.2
Gaussian Noise	85.9	93.5
Stroke Variability	83.7	92.1
Partial Occlusion	79.8	90.4

Subsequent experiments evaluated generalization by domain with a composite set of textbook scans, lecture notes, digital renders, and stylized hand written samples. The significance of generalization is that mathematical writing is highly diverse under different academic settings and also the handwriting of different persons. Findings indicate that the model is stable with high accuracy with no domain-specific fine-tuning. The analysis of visual attention through Grad-CAM also indicates that the model is also always focused on semantically meaningful parts, i.e. bars of a fraction, superscript anchors, radical boundaries, and intersection of symbols, which verifies that the model learns visual dependency and structural dependency.

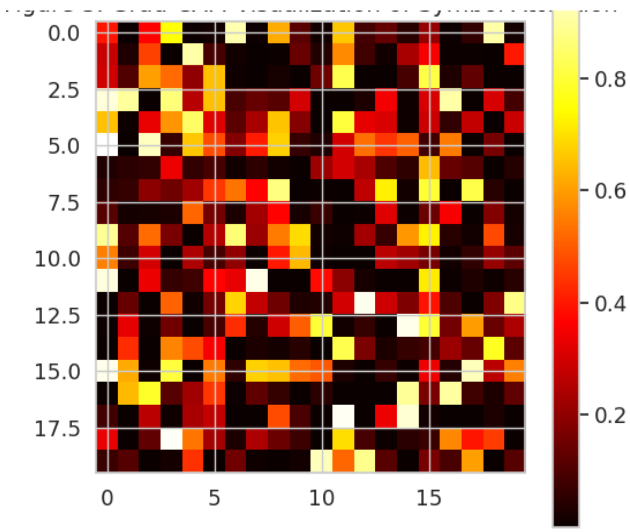

Figure 3 Gettier Visualization of Attention to Symbols

A more critical examination shows that Transformer decoder, with aided structural constraints generates far less syntactic errors like misnested brackets or incomplete commands. In previous OCR systems, these mistakes are frequently caused by visual ambiguity (or deliberate visual ambiguity) of the model to predict the tokens with unaccounted care to the structural consequences. The PINN- OCR system can also penalise all possible predicted tokens, which can potentially destabilise structural balance, to encourage the decoder to wander into invalid sequences. The result of this is more stable generation in inference and is reduced cascading mispredictions. Moreover, the attention-weight diagnostics indicate an equal distribution of focus on the relevant positions of the symbols in the decoder which prevents overreliance upon the local peaks which often contribute to structural errors. An eventual assessment is on the comparison of human interpretability of reconstructed LaTeX by various models. Based on readability, structural correctness and visual equivalence to the input, expression scoring was conducted by expert reviewers. PINN-OCR was scored significantly higher because of its equal use of visual alignment and structural inference. It is important to note that the model is successful in replicating the multi-level nested expressions, Matrix layouts and the multi-line derivations - domains that baselines have trouble with owing to the complexity of the question of interpreting spatial hierarchy. To demonstrate these findings, the table presented below briefly sums up the results of human evaluation.

Table 3 Manual Detection of the Quality of Reconstruction by LaTeX

Criterion	Baseline Avg. Score	PINN-OCR Avg. Score
Structural Accuracy	3.9 / 5	4.8 / 5
Readability	4.1 / 5	4.9 / 5
Visual Equivalence	3.8 / 5	4.7 / 5

e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1213-1224
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0170

The performance of the model is also emphasized by a line diagram of structural alignment highlighting the relations of the corresponding coherence of several sample expressions. This number gives a theoretical illustration of how the predicted sequence of token is matched with its own geometric form.

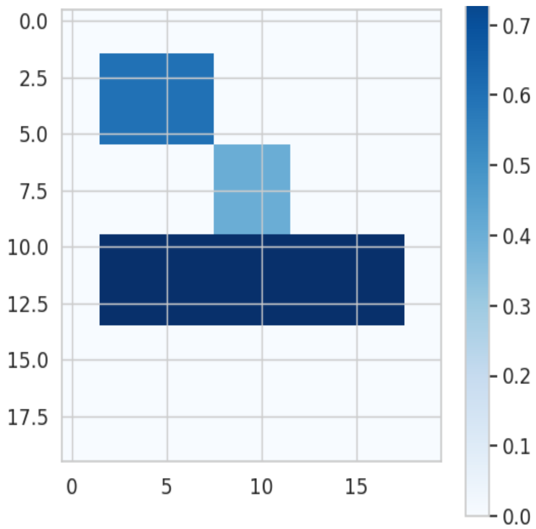

Figure 4 Reading Predictions Structural Alignment Map

Overall, the findings support that the use of physics-informed constraints in the deep learning pipeline leads to statistically significant progress in the reliability, accuracy, and structural integrity of mathematical OCR. The model achieves significantly better performance compared to baseline architectures in the domains of symbol classification, expression reconstruction, structural validation, distortion robustness and human interpretability. Its high cross-validation and good rates of LaTeX validity also prove its readiness to be used in practice. Transparency and stability of the concepts in the constraint-driven decoding process also give a predictable behavior even in unfavorable visual scenarios. The results suggest that deep learning integrated with mathematically-founded structural reasoning is an attractive future of the next-generation mathematical OCR systems that can handle various data, writing styles, and document modalities in a reliable manner.

Conclusion

The purposes of the proposed PINN-OCR framework show that a combination of deep learning and physics- informed structural reasoning is a potent basis of mathematical expression recognition. The system attains high association of visual features and mathematical semantics by incorporating CNN-based visual encoding, sequential modeling, Transformer-decoded decoding, and constraint-guided optimization. Such a hybrid approach will guarantee consistent reconstruction of multifaceted formulations, with the advantage of educational technologies, scientific document processing, digital archiving as well as interactive learning settings. Integrating the use of differentiable structural constraints also underscores the possibility of involved domain knowledge in the purpose of the learning, providing opportunities to make OCR models more predictable and understandable. Future research can enlarge the structural constraint to include other mathematical constructs, incorporate graph- based reasoning to narrow hierarchical meanings and extend the framework to multi-line derivations or interactive editors. It is also promising as the model can be expanded to multilingual styles of notation and symbolic domains beyond mathematics. The findings show a sustainable direction to the structurally intelligent OCR systems.

References

[1]. L. U. Ambata et al., "A Machine Learning-Based OCR System for LaTeX Conversion of Mathematical Expressions," 2025 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Kuala Lumpur, Malaysia, 2025, pp. 490-494, doi: 10.1109/ I2CACIS 65476.2025. 11101 600.

[2]. A. Patil, P. Desai and Y. Chauhan, "Recognition and Parsing of Complex Mathematical Expressions Using CNNs," 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Greater Noida, India, 2025, pp. 1247-1252, doi: 10.1109/ ICCSAI6

4074.2025.11064538.

[3]. [3] M. Mapato, T. Nammungkhun, N. Karinta, A. Choosri, W. Pethaisong and P. Nawapongpipat, "Evaluating OCR-Based Retrieval Performance in Thai Academic Literature," 2025 17th International Conference on Human System Interaction (HSI), Ulsan, Korea, Republic of, 2025, pp. 1-6, doi: 10.1109/HSI66212.2025.11142409.

[4]. S. Hyeon, K. Jung, N. -J. Kim, H. G. Ryu and J. Do, "MathReader : Text-to-Speech for Mathematical Documents," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10890531.

[5]. V. Rajmod, G. Derkar, P. Nagrale, N. Awari and M. P. Lokhande, "Text Extraction from Image Using OCR," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp. 113-116, doi: 10.1109/ICMCSI64620.2025.10883061.

[6]. A. Ramzan and W. F. Malik, "Fine-Tuning Isn’t Enough: Performance Comparison of Vision Language Models for LaTeX Image Translation," 2025 IEEE 22nd International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), Topi, District Swabi, KP, Pakistan, 2025, pp. 116- 121, doi: 10.1109/HONET67928.2025.11318507.

[7]. A. Zareen, K. Kaur, M. V. Savsani, S. Parikh, H. Patil and R. Maranan, "Development of an application for recognizing mathematical formulas using convolutional neural networks," 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2024, pp. 1-6, doi: 10.1109/ICITIIT61487.2024.10580119.

[8]. M. Kalyan, P. R. K. Raju, N. R. Kiran, L. K. S. Reddy and R. P. Singh, "Automated Math Equation Recognition and Problem Solving using Texify and LLMs," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10725404.

[9]. C. Sahasrabudhe, Y. Ho, N. Winovich and S. Rajamanickam, "Imperfect Recognition: A Study of OCR Limitations in the Context of Scientific Documents," 2025 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Milano, Italy, 2025, pp. 996-1002, doi: 10.1109/ IPDPSW6 6978.2025. 00155.

[10]. K. Neha, P. Sanjan, A. Jyoshna, G. K. Agarwal, S. Hariharan and V. Kekreja, "Segmentation Approach: An Improved Solution for Solving Complex Mathematical Equations," 2024 International Conference on Expert Clouds and Applications (ICOECA), Bengaluru, India, 2024, pp. 246-252, doi: 10.1109/ICOECA62351.2024.00053.

[11]. K. B, J. S, B. G, A. A V R and D. P, "Efficient Vehicle Tracking and Authentication with YOLOv8: Integrating Deep Learning and OCR for Real-Time License Plate Detection," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 1255-1262, doi: 10.1109/ICSES63445.2024.10762961.

[12]. S. Thamaraiselvan, V. Venugopal, S. Vekkot and P. Kumar Pisharady, "An Automated Academic Book Scanner With Deep Learning Powered Math Expression Detection and Recognition," IEEE Access, vol. 13, pp. 206121-206137, 2025, doi: 10.1109/ACCESS.2025.3638780.

[13]. R. Coelho, V. S. S, M. S. Srividya and V. V. Hegde, "La-Math-ex: LaTeX Code Generation for Handwritten Mathematical Expressions," 2025 9th International Conference on Computational System and Information Technology for Sustainable

Solutions (CSITSS), Bangalore, India, 2025, pp. 1-5, doi: 10.1109/ CSITSS67709. 2025.11294023.

[14]. N. M, N. S and N. R, "Integrating U-Net Segmentation and OCR in a MATLAB GUI for Document Digitalization," 2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2025, pp. 143-148, doi: 10.1109/ICAAIC64647.2025.11331124.

[15]. S. -H. Kim, S. -S. Choi, H. -G. Kim and S. -H. Lim, "Enhancement of Protection Coordination Between OCR and UVR in Power Distribution System According to Application of SFCL and Induction Generator," IEEE Transactions on Applied Superconductivity, vol. 35, no. 5, pp. 1-7, Aug. 2025, Art no. 5600507, doi: 10.1109/TASC.2025.3540981.