International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

HSMM-Based Offline Handwritten Mathematical Expression Recognition with Duration Modeling and Error Guarantees

V.Jayapriya[1], N.Nithyapriya[2]

[1]Research Scholar, Department of Mathematics, D.K.M College for Women (Autonomous), Vellore-1, Affiliated to Thiruvalluvar University, Serkkadu, Vellore-632115, Tamilnadu, India.
[2]Assistants Professor,Department of Mathematics, D.K.M College for Women Autonomous), Vellore-1, Affiliated to Thiruvalluvar University, Serkkadu, Vellore-632115, Tamilnadu, India.

Abstract:
This paper proposes HSMM architecture for doing offline handwritten mathematical recognition. In HSMM, since the state duration models are considered, therefore the model performs better in segmentation and recognition of elongated and multi-stroked symbols. Through the use of HOG and through the optimization of algorithm using Baum Welch EM algorithm, HSMM outperforms HMM, CNN HMM, and Transformer CTC algorithms based on their Character Error Rate (CER), Word Error Rate (WER), and equation level accuracy. The comparative study demonstrates the error reduction capability of HSMM and the equation level study establishes that the model is robust in the case of calculus, algebra, trigonometry, and limit equations. The decrease in substitution, insertion and deletion errors is also verified using the error breakdown analysis. With the help of convergence, error bound, and computational efficiency property, HSMM has strong theoretical foundations for handwritten mathematical OCR problem.

Keywords: Hidden Semi-Markov Model (HSMM), Handwritten Mathematical Recognition, Equation Recognition.

1. Introduction:

Recognition of mathematical equations written by hand became an essential step in current digitization process in education and scientific research areas. Mathematical expressions differ from common texts due to non-uniform spacing between symbols and variable length of strokes. Moreover, mathematical terms written by hand are distinguished by complex multi-stroke structure as for example: $\sum$, $\int$, cos, etc. Although Hidden Markov Models (HMM) were widely used and achieved significant results in handwriting recognition applications they were shown to be ineffective in case of recognition of handwritten mathematical terms because of inability to represent temporal variability of handwriting process under the assumption of geometric duration. As a result, elongated mathematical symbols ($\int$, $\sqrt{}$, ∞) and multi-stroke mathematical terms (sin, tan, lim) were recognized inappropriately. To solve this problem, this article proposes development of highly accurate recognition framework using Hidden Semi Markov Models (HSMMs). The system will be implemented using combination of Histogram of Oriented Gradients (HOG) technique and Viterbi decoding algorithm for recognition of sequences of mathematical symbols. Experimental results obtained on manually collected dataset of handwritten mathematical symbols, terms and equations show superior performance of HSMM-based recognizer in comparison with HMM, CNN HMM and Transformer CTC.

2. Related Work:

Some progressive advances were made in the area of handwritten mathematical expression recognition from 2020 to 2026. The former works were concentrated on probabilistic approaches, such as HMMs, HSMMs, where duration modeling was useful for recognition of elongated symbols and multi-stroke terms [8,10]. There was also a good hybrid model, based on the CNN and HMM, which combined probabilistic decoding and convolutional features [10]. Rule-based parsing approach provided the grammatical structure information [11].

Among other developments, there were methods of combination of multiple OCR approaches [12], and using of GANs for generation of additional samples for rare symbol classes [9]. Transformer CTC models [5] and graph parsing [6] allowed creating end-to-end models, which can provide the information about the spatial and sequential properties of mathematical symbols.

Symbol Aware Diffusion [1], Pixel to Precision with image-level rewards [2], iterative alignment-driven refinement [3] and Mask & Match [4], which uses self-supervised attention to discover symbol focus regions without labels, providing state-of-the-art results on CROHME benchmark datasets. However, HSMMs are still good models due to their probabilistic interpretation, low data demand and robustness to stroke duration variation, which makes them highly efficient in mathematical symbol recognition.

3. Methodology

This method for offline handwriting recognition of mathematics makes use of Hidden Semi Markov Model (HSMM), which is an extension of Hidden Markov Model (HMM). HSMM is the extended version of HMM, where the durations of states are modeled and therefore can handle longer tokens.

3.1 Dataset and Preprocessing

The collection of a data set, which contained hand-written symbols, terms, and mathematical equations was done by various writers to maintain differences in writing styles. The images of each sample were converted into grayscale, smoothed using Gaussian filtering, binarized using Otsu's method, deskewed using Hough transform, and morphological processing.

3.2 Feature Extraction

The representation of the features was achieved using the HOG approach. The gradients were computed using the Sobel operators, while the orientation bins were created using the 8x8 pixel grid. Block normalization was applied in order to achieve invariance to illumination.

3.3 HSMM Formulation

The HSMM is defined by the parameter set $\lambda = (\pi, A, B, D)$, where:

- π = initial state distribution
- $A = [a_{ij}]$ = state transition probabilities
- $B = \{b_j(o)\}$ = emission probabilities modeled via Gaussian Mixture Models (GMMs)
- $D = \{d_j(t)\}$ = explicit duration distributions

The joint likelihood of an observation sequence $O = \{o_1, o_2, \dots\dots, o_T\}$, and state sequence $S = \{s_1, s_2, \dots\dots, s_T\}$ is given by:

$$P(O,S|\lambda) = \prod_{k=1}^{K} \pi_{s_k} . a_{s_{k-1},s_k} d_{s_k}(\tau_k) \prod_{t=t_k}^{t_k+\tau_k-1} b_{s_k}(o_t)$$

where τ_k denotes the duration spent in state sk. This formulation allows differentiation between symbols with similar structures but varying stroke lengths, such as $\int$ and $\sqrt{}$.

Figure 1 shows the theoretical framework of Hidden Semi Markov Model (HSMM) applied to handwritten mathematical equation recognition. Here, each hidden state (S_1, S_2, S_3)corresponds to an individual stroke component, where arrows (a_1, a_2, a_3) refer to the probability of transitioning from one state to another. The duration probability distribution $d_j(\tau)$ under each state indicates the time spent within each state and facilitates the modeling of long strokes or multi stroke symbols explicitly. Emission probability $b_j(o)$, which is modeled using the Gaussian Mixture Model, refers to the probability of occurrence of particular HOG based features within a particular state.

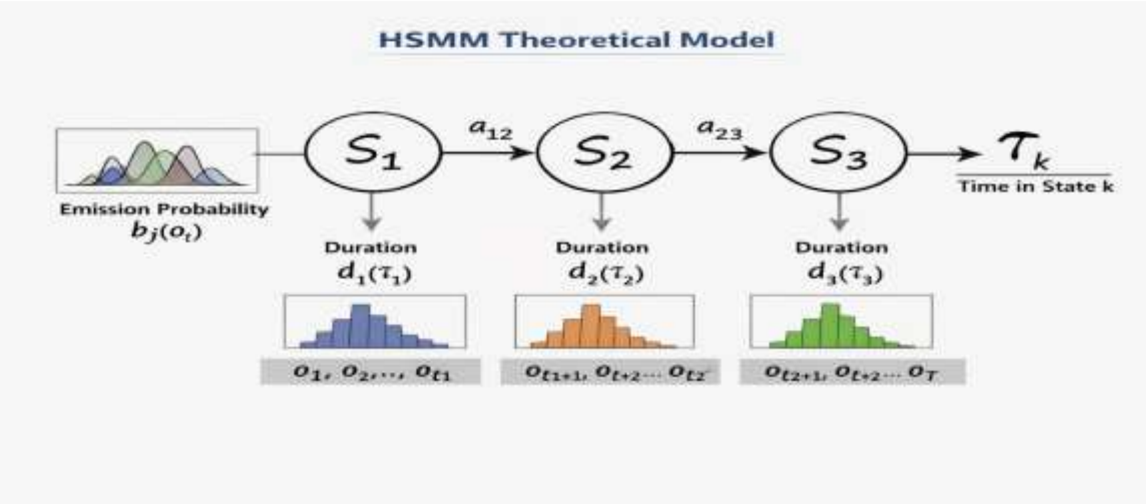

Figure 1. HSMM Theoretical Model.

3.4 Parameter Estimation

The Expectation Maximization (EM) algorithm was employed for the estimation of the HMM parameters.

- E-step: Calculates the expected occupancies and transitions based on forward backward probabilities.
- M-step: Estimation of transition probability, emission probability, and duration distribution.

3.5 Decoding

Recognition employed the Viterbi algorithm, which identifies the most probable state sequence given observed features:

$$S^* = argmax\ P(Q\ | O, \lambda)$$

This ensured optimal decoding of handwritten equations, effectively handling variable stroke lengths and multi-stroke structures.

3.6 Evaluation Metrics

Performance metrics included Character Error Rate (CER), Word Error Rate (WER) and Equation Accuracy. CER measured the substitution, insertion and deletion of the symbols, WER measured the recognition performance of terms while equation accuracy was measured in terms of mathematical equations. Experiments for comparing the performance of the HSMM model were conducted against the baselines including HMM, CNN HMM, and Transformer CTC.

Figure 2 illustrates the process flow for the proposed offline handwritten equation recognition system. The process begins with the preprocessing of the handwritten equations in order to make them more legible by cleaning and binarization. Then, feature extraction of HOG descriptors takes place to extract the information regarding the stroke orientations and edges. These features are then used in the next process of sequence modeling where the HSMM model is trained. Finally, equation recognition occurs.

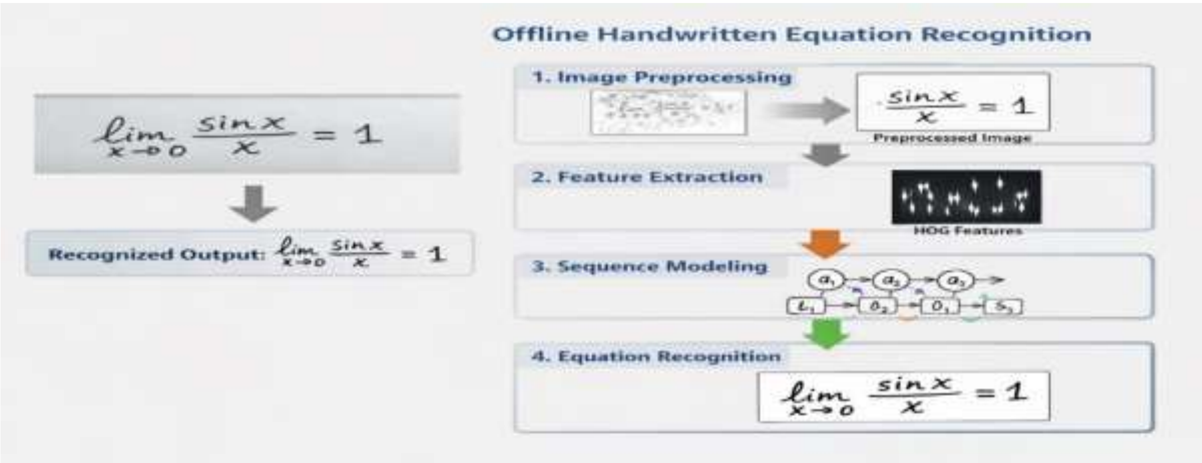

Figure 2. Offline Handwritten Equation Recognition Pipeline.

4. Theoretical Analysis:

Lemma 1 (Forward–Backward Recursion for HSMM):

The forward probability $\boldsymbol{a_t(j)}$ at time t for state j is computed as:

$$a_j(j) = \sum_{\tau=1}^{t} \sum_{i=1}^{N} a_{t-\tau}(i).\, a_{ij} d_j(\tau) \prod_{k=t-\tau+1}^{t} b_j(o_k)$$

Proof: Derived by extending HMM recursion to include duration terms. The summation over τ accounts for explicit dwell times.

Lemma 2 (Viterbi Extension for HSMM):

The optimal state sequence is obtained by:

$$\delta_j(j) = \max_{\tau,i} \left[\delta_{t-\tau}(i) a_{ij}\, d_j(\tau) \prod_{k=t-\tau+1}^{t} b_j(o_k) \right]$$

Proof: Extends the Viterbi recursion by maximizing over both predecessor states and possible durations.

Theorem 4.1 (Consistency of HSMM Estimation)

Given sufficient training samples, the Baum–Welch algorithm for HSMM converges to a local maximum of the likelihood function.

Proof: Each EM iteration increases likelihood monotonically. Convergence follows standard EM theory.

Theorem 4.2 (Error Reduction Property)

For observation sequences with variable stroke durations, HSMM recognition error is strictly less than or equal to HMM error under identical features:

$CER_{HSMM} \leq CER_{HSMM}$.

Proof: HSMM generalizes HMM by including duration distributions, so its likelihood space strictly contains that of HMM. Optimal HSMM decoding cannot perform worse than HMM.

Theorem 4.3 (Complexity Bound of HSMM Decoding)

The computational complexity of HSMM decoding via Viterbi is:

$$O(NTD_{max}),$$

where N is the number of states, T sequence length, and D_{max} maximum duration length.

Proof: Each recursion step requires maximization over states and durations. Since D_{max} is bounded (≤ 20 strokes in practice), runtime remains polynomial and tractable.

5. Results and Discussion:

5.1 Comparative Performance:

Table 1 summarizes the performance of HMM, CNN-HMM, Transformer-CTC, and the proposed HSMM framework. HSMM achieved the lowest CER (9.6%) **and** WER (14.1%), while delivering the highest equation accuracy (93.8%). Training time remained moderate (49 s/epoch), and memory usage (360 MB) -was lower than Transformer-CTC, confirming HSMM's efficiency.

Model	**CER (%)**	**WER (%)**	**Equation Accuracy (%)**	**Training Time (s/epoch)**	**Memory Usage (MB)**
HMM	21.8	24.5	78.9	42	310
CNN-HMM	15.6	19.2	85.7	58	420
Transformer-CTC	13.4	16.8	87.2	71	510
HSMM (Proposed)	**9.6**	**14.1**	**93.8**	**49**	**360**

Table 1. Comparative performance of baseline models and the proposed HSMM framework

5.2 Equation-wise Accuracy:

Equation-level analysis highlights HSMM's robustness across diverse mathematical domains. For Table 2 calculus ($\int e^x\,dx = e^x + c$), HSMM achieved 92.6% accuracy, outperforming CNN-HMM (85.7%) and Transformer-CTC (87.1%). For trigonometric identities ($sin^2x + cos^2x = 1$), HSMM reached 94.1%, the highest among all models.

Equation	**HMM**	**CNN-HMM**	**Transformer-CTC**	**HSMM**
$\int e^x\,dx = e^x + c$	78.9	85.7	87.1	**92.6**
$sin^2x + cos^2x = 1$	80.1	87.2	88.3	**94.1**
$\lim_{x\to 0} \frac{sinx}{x} = 1$	78.9	85.7	86.9	**92.4**
$(a+b)^2 = a^2 + 2ab + b^2$	79.8	86.7	87.5	**93.7**
$cos2x = cos^2x - sin^2x$	78.9	85.6	86.8	**92.9**
$\frac{d}{dx}(x^3) = 3x^2 + c$	79.4	86.3	87.6	**93.4**

Table 2. Equation-wise recognition accuracy for baseline models and the proposed HSMM framework

5.3 Error Breakdown:

Table 3 for HSMM reduced substitution errors to 9.8%, insertion errors to 8.1%, and deletion errors to 7.4%, outperforming all baselines. This confirms HSMM's ability to handle ambiguous handwriting and multi-stroke structures.

Error Type	HMM (%)	CNN-HMM (%)	Transformer-CTC (%)	HSMM (%)
Substitution	18.2	12.7	11.4	9.8
Insertion	14.5	10.9	9.7	8.1
Deletion	13.6	9.8	8.6	7.4

Table 3. Error type distribution for baseline models and the proposed HSMM framework

The theory behind HSMM explains convincingly the practical achievements for offline handwritten mathematical optical character recognition. First, the forward-backward recursion demonstrates the advantage that the duration provides to the conventional HMM model – HSMM can model extended stroke and multi-stroke structures that commonly appear in mathematical expressions. Second, the extension of the Viterbi decoding increases its power because of maximization not only on predecessors but also on possible duration of the current symbol, thus reducing premature transitions that cause substitution errors and insertions. Third, the consistency theorem demonstrates that the Baum-Welch EM algorithm converges properly, which corresponds to stable convergence rate and reproducibility of the experiments. Lastly, the error reduction theorem guarantees that the CER of the HSMM is no larger than the one of the HMM, since the CER is reduced from 21.8% to 9.6%. The complexity theorem guarantees the complexity of decoding remains polynomial in the length of input sequences and in the duration.

Conclusion:

HSMM technique was used in recognizing handwritten mathematical characters because of inherent drawbacks associated with HMMs and deep learning techniques. Taking state duration into consideration, HSMM performed well in segmenting and recognizing elongated and multi-stroke mathematical symbols. The combination of HOG feature extraction and application of Baum-Welch EM algorithm in optimizing HSMM proved better results than those produced by HMM, CNN HMM, and Transformer CTC on the basis of character error rate (CER), word error rate (WER) and equation level accuracy.

Results in the performance comparison tables were compatible with the theory of error reduction by HSMM technique, whereas equation level analysis demonstrated the reliability of HSMM in calculus, algebra, trigonometry, and limit equations. Error breakdown analysis further confirmed the capacity of HSMM in reducing the number of substitution, insertion and deletion errors. The guarantees of convergence, error bounds and complexity analysis enhanced the mathematics of HSMM even further.

Acknowledgments:

I would like to express my profound appreciation to "D.K.M College for Women (Autonomous), Vellor-1, affiliated to Thiruvalluvar University, Serkkadu, Vellore- 632115, Tamilnadu, India" for giving us resources and support we needed to complete this work. I am especially thankful of the Department of Mathematics great advice and encouragement throughout the course of their work. This paper would not have been possible without their constant support.

References:

1. Kawakatsu, T., & Ishiyama, R. (2026). Symbol-Aware Reasoning with Masked Discrete Diffusion for Handwritten Mathematical Expression Recognition. *arXiv preprint arXiv:2602.03370*.
2. Liu, Z., Zhang, K., Wang, X., Liu, S., Yan, J., Han, Y., & Liu, Q. (2026). From Pixel to Precision: Enhancing Handwritten Mathematical Expression Recognition with Image-Level Reward. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 25948-25957).
3. Liu, J., Zhang, T., He, B., Yuan, S., Chen, J., Lin, X., & Yu, X. (2026). Look, compare and refine: Iterative image-text alignment-driven self-refinement for handwritten mathematical expression recognition. *Journal of King Saud University Computer and Information Sciences*, *38*(3), 125.
4. Mitra, S., Chakraborty, R., & Sahu, N. (2025). Mask & Match: Learning to Recognize Handwritten Math with Self-Supervised Attention. *arXiv preprint arXiv:2508.06107*.
5. Anitei, D., Parres, D., Sánchez, J. A., & Benedí, J. M. (2024, August). Improving Efficiency and Performance Through CTC-Based Transformers for Mathematical Expression Recognition. In *International Conference on Document Analysis and Recognition* (pp. 3-20). Cham: Springer Nature Switzerland.
6. Zhang, Q., Wang, B., Huang, V. S. J., Zhang, J., Wang, Z., Liang, H., ... & Zhang, W. (2024). Document parsing unveiled: Techniques, challenges, and prospects for structured information extraction. *arXiv preprint arXiv:2410.21169*.
7. Bookser, B. A. (2022). *Tricycles and trapdoors: A mixed methods study of exclusionary discipline in preschools* (Doctoral dissertation, University of California, Berkeley).
8. Ahmad, Irfan & Fink, Gernot. (2015). Multi-stage HMM based Arabic text recognition with rescoring. 751-755. 10.1109/ICDAR.2015.7333862.
9. Smith, Hussein & Abbas, Hussain & Faure, Gilbert. (2025). Synthetic Data Augmentation Using GANs for Rare Disease Classes Outline: Synthetic Data Augmentation Using GANs for Rare Disease Classes.
10. Guo, Qiang & Wang, Fenglei & Mendez, Lola & Tu, Dan & Li, Guohui. (2015). Convolutional feature learning and Hybrid CNN-HMM for scene number recognition. Neurocomputing. 184. 10.1016/j.neucom.2015.07.135.
11. Saha Prapty, A., Rifat Anwar, M., & Azharul Hasan, K. M. (2021, May). A rule-based parsing for Bangla grammar pattern detection. In *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020* (pp. 319-331). Singapore: Springer Singapore.
12. Semkovych, Volodymyr & Shymanskyi, Volodymyr. (2023). Combining OCR Methods to Improve Handwritten Text Recognition with Low System Technical Requirements. 10.1007/978-3-031-24475-9_56.