JSCDM

JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 7 No. 2 (2026) 45-57
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Leveraging Cellular Automata for Strip-Cut Document Reconstruction: A Path Travelling Salesman Problem Formulation

Abdallah S. Hyassat[1,2]*, Insaf Kraidia[3], Mosleh M. Abualhaj[4], Iyas Qaddara[5], Abdel Latif Abu Dalhoum[2]

[1] *Data Science and Artificial Intelligence Department, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, 19328, JORDAN*
[2] *Computer Science Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, 11942, JORDAN*
[3] *Software Engineering Department, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, 19328, JORDAN*
[4] *Networks and Cybersecurity Department, Al-Ahliyya Amman University, Amman, 19328, JORDAN*
[5] *Computer Science Department, Al-Ahliyya Amman University, Amman, 19328, JORDAN*

*Corresponding Author: A.hyassat@ammanu.edu.jo
DOI: https://doi.org/10.30880/jscdm.2026.07.02.004

Article Info

Received: 11 November 2025
Accepted: 21 January 2026
Available online: 30 June 2026

Keywords

Shredded document reconstruction, Cellular Automata, strip-cut documents, Travelling Salesman Problem

Abstract

Reconstructing shredded documents is a long-standing problem in information security, forensics, and archival recovery. This work tackles the Reconstruction of Shredded Documents Problem (RoSDP) by formulating it as a path Travelling Salesman Problem (TSP) and solving it using Cellular Automata (CA). The proposed method is framed as an intelligent control system, where the CA operates as a decision-making engine that determines optimal sequential actions for strip arrangement, similar to automated control strategies employed in robotics. The pipeline includes dataset construction, edge-based dissimilarity computation, an order-respecting 2D coordinate mapping that overcomes non-metric multidimensional scaling (NMDS) failures, CA-driven search to recover the left-to-right shred order, and deterministic page reassembly. Evaluated on 120 pages stratified by content density and shredded into 20, 25, and 30 strips. The method delivers a mean page-level reconstruction accuracy of 70.74%, achieves perfect 100% recovery on many pages, and runs in approximately 30–55 seconds per page on a single CPU. Analysis shows errors stem primarily from the boundary dissimilarity cost function rather than the CA mechanism itself. Compared with a closely related optimisation-based baseline, the proposed approach yields higher overall accuracy. These results position CA as a practical, reproducible, and competitive intelligent-control framework for strip-cut reconstruction and motivate future improvements to the cost function and extensions to cross-cut and hand-torn documents.

1. Introduction

Reconstruction of broken objects to recover the original artefact is a complex and time-consuming process. It plays an important role in several domains such as archaeology, art restoration, forensics, and investigation science [1]. Generally, the shredded object may be an image, a document, or anything else. In the document domain, paper shredders fragment pages into pieces that vary in shape, size, and density. There are three main categories of shreds, and these categories are: strip-cut, cross-cut, and hand-shredded or torn. In strip-cut, the object is cut into a long vertical strip. In cross-cut, they are cut both horizontally and vertically into small rectangles. In hand-shredded, fragments are irregular in shape [2]. Fig. 1 illustrates these shredding categories. Although Shredders break each page into a large number of visually similar strips, these strips retain informative edge patterns. If reinstated, the recovered content can have a major effect on criminal inquiry, military intelligence, and historical study [3]. The present study, therefore, focuses on demonstrating that Cellular Automata (CA) can solve the Reconstruction of Shredded Documents Problem (RoSDP) for strip-cut pages.

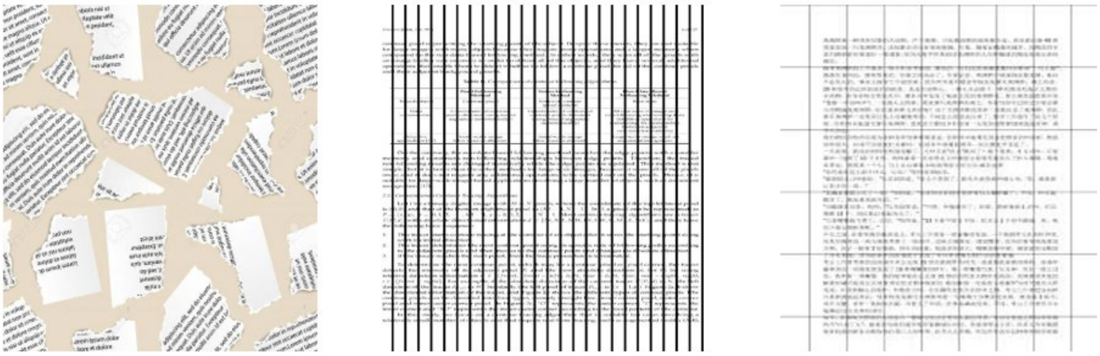


Fig. 1 *Examples of different document shredding types: (a) Manually torn document; (b) Strip-cut shredded document; and (c) Cross-cut shredded document*

Computer-aided systems are now an integral part of core workflows across many domains, including image analysis [3], supply chain [4], medical applications [5], optimisation-based decision support [6,7], and security [8]. By the same rationale, automating shredded-document reconstruction replaces slow, error-prone manual assembly with faster, more consistent results. Accordingly, the main objective of this study is to present an automated method to solve RoSDP using CA by mapping RoSDP to a path Travelling Salesman Problem (TSP) instance and then applying CA to that instance; once the path-TSP is solved, RoSDP is solved accordingly. The aim is to develop a method that is faster, more efficient, and more practical than alternative techniques for reconstructing shredded documents. The contributions of this paper can be summarised as follows:

- Casts RoSDP as a path TSP over directed boundary dissimilarities and shows how solving this path suffices to reconstruct a shredded page.
- Introduces an order-preserving 2D embedding method that enforces left-to-right progression while remaining faithful to local edge costs, overcoming non-metric MDS failures.
- Designs a deterministic CA traversal that functions as an intelligent-control loop to recover the strip sequence efficiently and reproducibly.

2. Background

This section outlines the theoretical foundations that underpin our proposed method. We briefly review key concepts, including CA, the TSP, and NMDS, as they form the basis for formulating and solving the RoSDP.

2.1 Cellular Automata (CA)

CA are discrete computational systems that consist of a vast number of simple cells through a grid. A cell has a fixed state; the cells update their state simultaneously and with discrete time steps using a local transition rule based only on the state of the cell itself and that of its neighbours. CA is effective for modelling systems in which complex global behaviour emerges from simple, local interactions [9]. Owing to this property, CA has been widely used to model and simulate complex phenomena across science and engineering [10], and have found practical applications in areas such as image generation and scrambling [11], pseudo-random number generation [12], pattern recognition [13], traffic routing [14], and games [15]. Cellular automata are computational models that have the following characteristics:

- The cells live on a grid
- Each cell has a state
- Each cell has a neighbourhood.

CA can be defined on grids of different dimensionalities; most commonly are one-dimensional (1D) and two-dimensional (2D) [16]. In 1D CA, the next state of each cell depends on its current state and those of its neighbours along the line [17]. Elementary Cellular Automata (ECA) are the simplest 1D CA, with binary states {0,1} [18]. In 2D CA, an initial state is assigned to each cell on a 2D grid, and, at each iteration, cells update according to a transition function that depends on the previous states of the cell and its neighbours [19]. Two commonly used neighbourhood structures in 2D CA are the von Neumann neighbourhood and the Moore neighbourhood, which are illustrated in fig. 2, which includes both orthogonal and diagonal neighbours. In this paper, a restricted form of the Moore neighbourhood is employed by pruning the left-hand side of each expansion ring. This restriction follows naturally from the adopted 2D placement strategy, which enforces a nondecreasing progression along the x-axis.

D (x-1,y-1)	*D* (x,y-1)	*D* (x+1,y-1)
D (x-1,y)	*P* (x,y)	*D* (x+1,y)
D (x-1,y+1)	*D* (x,y+1)	*D* (x+1,y+1)

Fig. 2 *Moore neighbourhood*

2.2 Travelling Salesman Problem (TSP)

TSP is a fundamental and significant optimization problem in transportation applications, this means that there is a salesman who must visit each city once and never go back to that city again and return to the first city he visited at the begging, whereas this route must be the shortest route available, taking into account that travelling from a city to another city has a cost (distance) of travelling for each city [20]. This problem belongs to the class of NP-hard problems, which requires an exponential computation time. The main idea of exponential problems is that they are ideal for a small number of instances. However, if the input size increases, the time needed to solve the problems significantly increases. Therefore, a large number of heuristic algorithms were developed to discover near optimal or optimal solutions effectively [21,22].

In this paper, the reconstruction problem is formulated as a path-TSP rather than a traditional cycle TSP. This choice is motivated by the structural properties of shredded documents, where the leftmost and rightmost strips act as fixed terminals and do not form a natural closed loop. Unlike the cyclic TSP, the path-TSP formulation allows an open tour that starts from the leftmost strip and ends at the rightmost one. This formulation better reflects the reconstruction process and avoids imposing unnecessary constraints that are incompatible with the problem.

2.3 Non-Metric Multidimensional Scaling (NMDS)

Also known as Non-Classical Multidimensional Scaling. NMDS aims to represent the original position of the data in multidimensional space as accurately as possible using a small number of dimensions that can be easily plotted and visualised. Both Metric Multidimensional Scaling (MDS) and NMDS do the same job, which is taking a matrix as an input and producing a configuration of points for the elements in the matrix. The main difference between them is that in MDS, values in the input matrix should represent real distances between elements, whereas NMDS takes a (dis)similarity matrix as input, and these (dis)similarity values don't represent real distances [23].

In this paper, NMDS is used to approximate the relative position of the strips according to their relative dissimilarity with each other. Since these dissimilarities are not the true metric distances but represent the relative similarity using the information at boundaries, classical multidimensional scaling is poorly adapted to that situation. NMDS, in its turn, relies on the rank-based order of dissimilarities as opposed to their absolute values, which makes it suitable for allowing the strips to be arranged in a way that makes their following navigation and reconstruction easier.

3. Literature Review

Early efforts on reassembling shredded or torn pages relied on contour continuity and shape cues. Using contour maps, hand-torn pages of eight fragments were reconstructed with an average joining time of about 0.5 seconds [1]. Feature-matching pipelines subsequently targeted hand-torn documents; for example, a method for 3–16 fragments ranging from 1×1 cm to 5×5 cm was described, but performance dropped as fragment count grew, reaching roughly 52% at 15 pieces [24]. A more elaborate framework digitised each fragment, extracted local shape- and content-based features, and used an SVM to identify reliable support-point matches, followed by geometric and content constraints to score group alignments; tested on 120 pages ripped by hand into 16 parts, the system reassembled pages with up to 32 fragments [25].

For strip-cut text documents, an architecture was evaluated on documents shredded into five pieces; handwritten material achieved a precision near 1.0, whereas colour-layout pages performed poorly [26]. An image-based technique that prints the original on A4 and shreds it into 21 pieces was presented [27]. Mapping strip-cut reconstruction to combinatorial optimization has proven effective: RSSTD was formulated as a TSP, solved with Lin–Kernighan heuristics and aided by user-guided Variable Neighborhood Search [28]. Color-space matching with HSV/RGB features and a nearest-neighbor Euclidean distance, combined with a winner-takes-all decision, was also explored [29]. Domain-specific evaluations further documented practical preprocessing and featurization choices and demonstrated applicability to non-Latin scripts and governmental records, indicating generalization beyond English texts [30].

Cross-cut reconstruction is more challenging due to two-dimensional adjacency. Higher accuracy and lower runtime than a TSP-only baseline were reported [31, 32]. Near-perfect precision was achieved when the number of fragments was below 55, degrading as fragments increased [33]. A multi-photo assembly approach based on character features was proposed [34]. A splicing-driven memetic algorithm (SD-MA) exploited adjacency information through crossover and mutation, outperforming well-known baselines such as SD-GA and HV2 across 30 cases from 10 multilingual documents in both exactness and convergence time [35]. A high-accuracy solution combined divide-and-conquer with a reduction from cross-cut to multiple Reconstruction of Strip Shredded Text Documents instances, which were then solved using an Ant-Colony Optimization Algorithm. The experiments were carried out on 10 documents, with five documents being in Chinese and the other five in English, and the accuracy rates stood at about 95% and 85% of the documents, respectively, with this difference being partially ascribed to boundary characteristics and font uniformity differences [36]. Extending this line, clustering was improved via horizontal projection and a constrained seed K-means algorithm [37]. More recently, for cross-cut documents, an extreme learning machine–based compatibility model aggregated consensus information from edge and layout signals and improved matching accuracy and robustness relative to classical baselines [38].

4. Methodology

Our pipeline proceeds in five stages, as illustrated in fig. 3. We first build our dataset; next, we formulate the reconstruction problem as a path-TSP using dissimilarities computed between pairs of edges; next, we find the points that represent the shreds. A CA is then used to solve the TSP. Lastly, the torn-up document is reconstructed according to the received sequence.

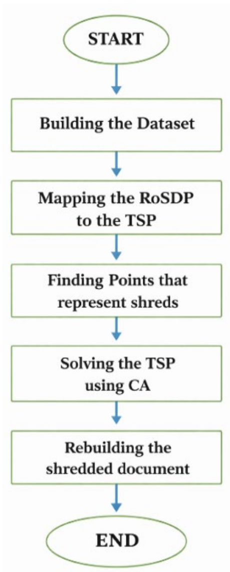
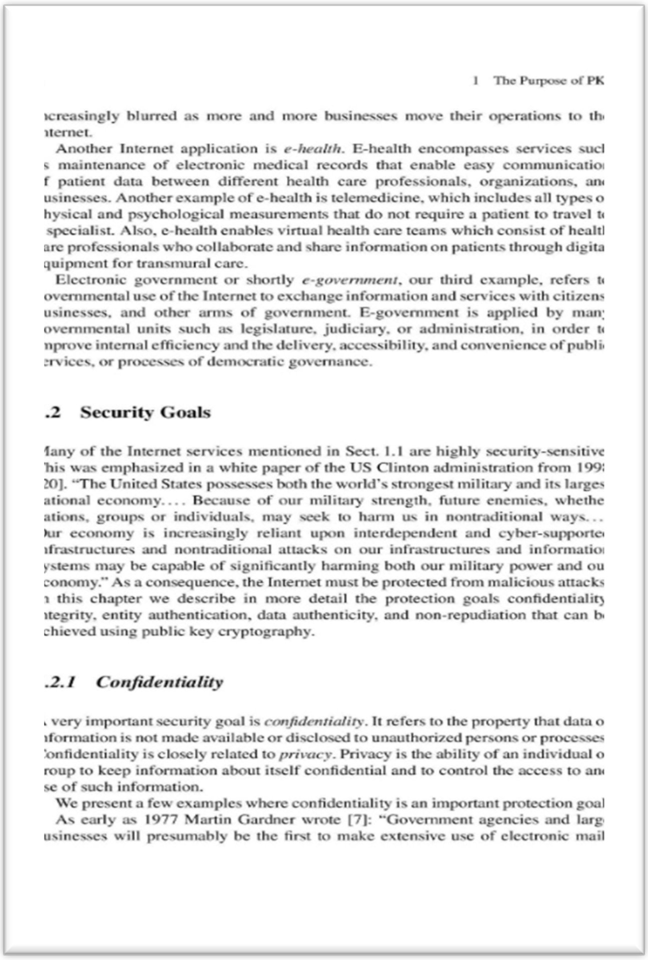


Fig. 3 *End-to-end workflow of the proposed method*

4.1 Dataset construction and virtual shredding

The corpus is drawn from the published book “Introduction to Public Key Infrastructures”, from which 120 pages are selected to provide a consistent and citable benchmark. All pages are transformed into .jpg files and loaded in MATLAB as $N \times M$ grayscale images where intensities fall in [0,255]. A median filter is used to normalize noise and eliminate uniform left-right margins so that comparisons of edges involve content, not blank borders. An attempt to simulate strip-cut shredding in a predictable environment is to control each normalized page by dividing it up into n-wide vertical strips, where $n \in \{20, 25, 30\}$. On every run, strip identifiers are randomly permuted to prevent order leakage. The boundary anchors are then identified by scanning the five outermost columns for near-uniform white, and the detected leftmost and rightmost strips are fixed at positions 1 and n, respectively. Fig. 4 shows an example original paper, and Fig. 5 shows its shredded representation.



1 The Purpose of PK

ıcreasingly blurred as more and more businesses move their operations to th ıternet.

Another Internet application is *e-health*. E-health encompasses services sucl s maintenance of electronic medical records that enable easy communicatio f patient data between different health care professionals, organizations, an usinesses. Another example of e-health is telemedicine, which includes all types o hysical and psychological measurements that do not require a patient to travel t specialist. Also, e-health enables virtual health care teams which consist of healtl are professionals who collaborate and share information on patients through digita quipment for transmural care.

Electronic government or shortly *e-government*, our third example, refers t overnmental use of the Internet to exchange information and services with citizens usinesses, and other arms of government. E-government is applied by man overnmental units such as legislature, judiciary, or administration, in order t nprove internal efficiency and the delivery, accessibility, and convenience of publi ervices, or processes of democratic governance.

.2 Security Goals

Iany of the Internet services mentioned in Sect. 1.1 are highly security-sensitive 'his was emphasized in a white paper of the US Clinton administration from 199 20]. “The United States possesses both the world’s strongest military and its larges ational economy.... Because of our military strength, future enemies, whethe ations, groups or individuals, may seek to harm us in nontraditional ways...)ur economy is increasingly reliant upon interdependent and cyber-supporte ıfrastructures and nontraditional attacks on our infrastructures and informatio ystems may be capable of significantly harming both our military power and ou conomy.” As a consequence, the Internet must be protected from malicious attacks ı this chapter we describe in more detail the protection goals confidentiality ıtegrity, entity authentication, data authenticity, and non-repudiation that can b chieved using public key cryptography.

.2.1 Confidentiality

. very important security goal is *confidentiality*. It refers to the property that data o ıformation is not made available or disclosed to unauthorized persons or processes 'onfidentiality is closely related to *privacy*. Privacy is the ability of an individual o roup to keep information about itself confidential and to control the access to an se of such information.

We present a few examples where confidentiality is an important protection goal As early as 1977 Martin Gardner wrote [7]: “Government agencies and larg usinesses will presumably be the first to make extensive use of electronic mail

Fig. 4 *Example original page before shredding*

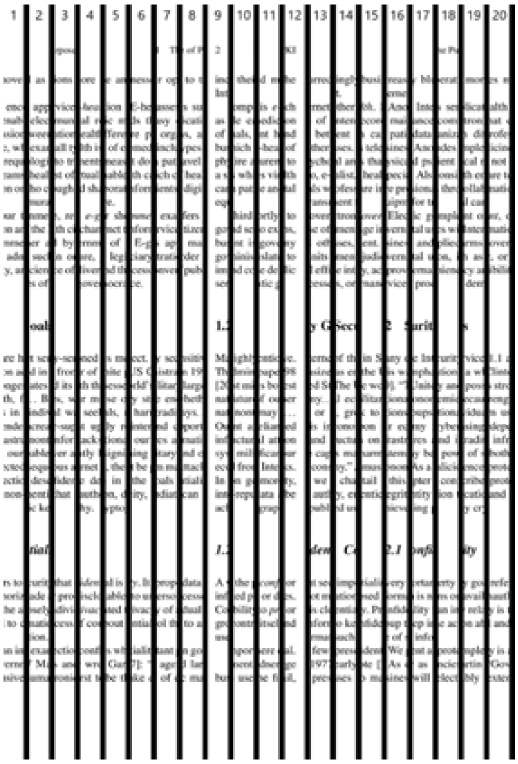


Fig. 5 *Virtual strip-cut representation*

4.2 Mapping RoSDP to Path-TSP

In this paper, we formulate the RoSDP as a path-TSP. The idea is to recover the correct left-right sequence of strips based on their vertical boundaries. The strips are represented as nodes in a directed graph, and the cost associated with transiting between two nodes is the visual compatibility between the two strips. An ordered pair of strips (i, j) should have a directed dissimilarity measure $D(i, j)$, which is used to measure how well the strip j can be directly laid to the right of the strip i. This measure is calculated by comparing the right boundary of the strip i and the left boundary of the strip j using a row-wise absolute difference as shown in Equation 1.

$$D(i, j) = \sum_{r=1}^{n} |strip_j(r, 1) - strip_i\,(r, C)| \quad (1)$$

Where N denotes the strip height, C is the strip width, $strip_i\ (r, C)$ represents the pixel value at row r on the right edge of the strip i, and $strip_j\ (r, 1)$ denotes the corresponding pixel on the left edge of the strip j. A lower value of $D(i, j)$ indicates stronger visual continuity and thus a higher likelihood that j follows i in the correct reconstruction.

The boundaries of the leftmost and the rightmost strips are known in advance with the help of the analysis of the boundary content. The strip on the left does not visually contain anything along its leftward edge, and the strip on the right, as well, does not contain anything along its rightward edge, allowing both to be viewed as fixed endpoints during reconstruction. A square dissimilarity matrix of size $(n - 1) \times (n - 1)$ is therefore constructed. The reason is that the reconstruction process starts with the identified leftmost strip and only transitions between the remaining strips are taken into account, while the rightmost strip is added at the end of the process when the optimal ordering is obtained.

Following the definition of the boundary dissimilarity, the reconstruction task can be directly formulated as a path-TSP problem. In contrast with the classical TSP, which aims to have a closed cycle, the current formulation is that of a path-TSP, as the document has terminal points. The traversal, therefore, starts from the detected leftmost strip and finishes with the rightmost strip attached after the optimal ordering has been acquired. Here, Figure 6 shows one of the examples of the dissimilarity matrix calculated by applying the boundary cost function.

0	1	2	3	4	5	6	7	8	9
1	0	21160	21586	17445	20413	21208	21335	21246	17224
2	93628	0	21507	13470	20781	20387	18975	17985	16321
3	94352	21265	0	14569	21472	18888	18109	18159	15894
4	99045	26621	23605	0	22100	25914	24056	24159	16222
5	119061	45813	44707	32560	0	45501	45018	45004	19454
6	75886	3550	2852	3307	3847	0	2385	2193	3488
7	78534	6260	5485	2940	6645	3322	0	960	5403
8	75343	2912	2241	2310	3290	1172	1265	0	2689
9	108455	35989	32913	24892	29748	34201	32861	34482	0

Fig. 6 *Dissimilarity matrix*

4.3 Monotone 2D Placement of Strips

To preserve the left-to-right order required to correctly reconstruct the documents, a constrained 2D placement strategy is used. The goal is to create virtual positions for the strips where these positions satisfy the ordering constraint and the local dissimilarities of the strips' boundary defined by the dissimilarity matrix. The initial strip is placed at the position $(0, 0)$. The position of the strip $i + 1$ is calculated relative to the previously placed strip i. The following point is restricted to lie on a circle with its center at (x_i, y_i) with radius $D(i, i + 1)$. To ensure the left-to-right motion, only the right half of this circle is considered valid. The relative vertical movement between two adjacent strips is governed by a displacement parameter Δy, drawn uniformly from the range $[-\mathrm{D}, \mathrm{D}]$, which controls the allowable vertical variation during placement. After a value of Δy is chosen, the vertical coordinate of the next strip is calculated using Equation (2). Equation (3) is then used to obtain the corresponding x-coordinate.

$$y_{i+1} = y_i + \Delta y \tag{2}$$

$$x_{i+1} = x_i + \sqrt{D^2 - \Delta \mathrm{y}^2} \tag{3}$$

Where $(x_i,\ y_i)$ and $(x_{i+1},\ y_{i+1})$ denote the coordinates of the current and subsequent strips, respectively, Δy represents their vertical displacement, and D represents the boundary dissimilarity between two consecutive strips and is used as the radius in the circular placement constraint. To further avoid invalid placements, the candidate point is rejected in case it violates the ordering constraint. Specifically, in cases where the value of x is less than x_i or if its distance to a previously placed strip is smaller than the distance to its immediate predecessor, that candidate point is eliminated, and a new Δy is selected. Such an acceptance criterion makes sure that every new strip moves forward in space and never intersects or moves backward of strips placed before it. Figure 7 depicts this positioning strategy. The blue arc represents the set of valid positions of the strip $i\ + 1$, which will be the right semicircle of radius $D(i, i + 1)$ whose center is at the location of the strip i. The points that are on such an arc are the only points that qualify as viable candidates because they ensure that progress is made in the rightward direction. Any point not in this region is discarded, as it would violate the left-to-right ordering constraint.

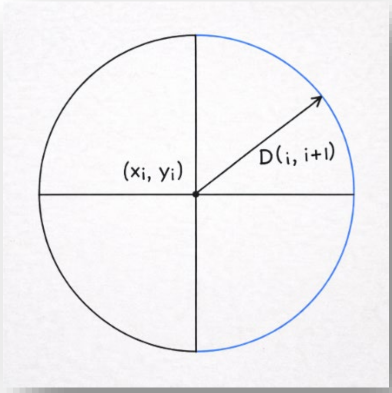


Fig. 7 *Valid region for point i+1*

After laying all strips, the complete layout is translated to ensure that all the coordinates are positive and fall within the border of the CA grid. The resulting arrangement preserves the sequential ordering of strips while approximately maintaining their pairwise boundary dissimilarities, and it serves as the initial state for the subsequent CA traversal.

4.4 Ordering with a Cellular Automata

Given the order-preserving coordinates, we cast sequencing as a deterministic traversal on a two-dimensional CA. The CA grid is sized to the maximum placed coordinates plus a small padding so that all valid locations lie strictly inside the boundary. All cells are initialized to state 0; cells that coincide with strip coordinates are set to state 10 and treated as shred cells (SCs), fig. 8 shows this initialization.

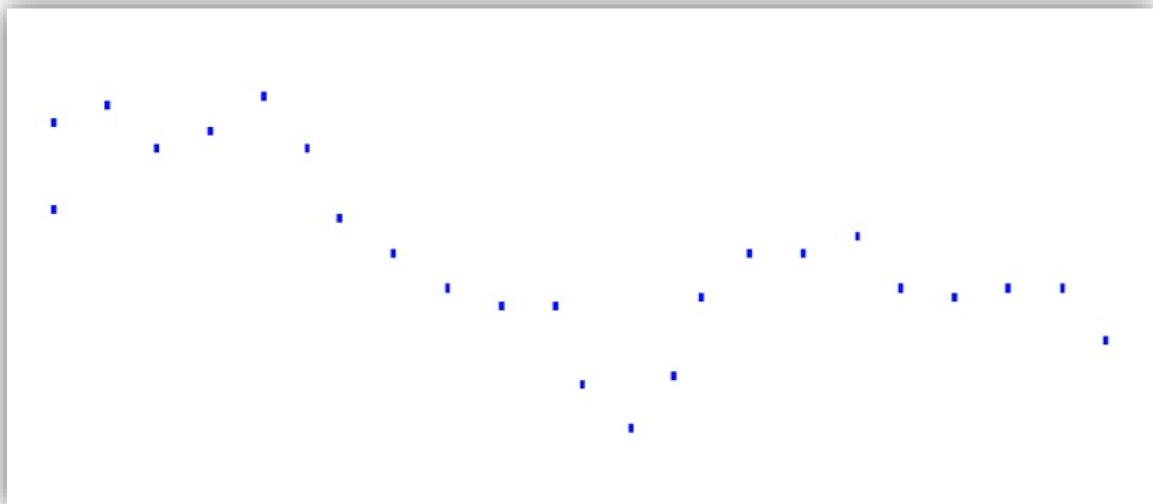


Fig. 8 *Order-preserving embedding*

The traversal starts at the SC corresponding to strip 1 (the left anchor), the CA expands as concentric rings using the constrained Moore neighborhood. At ring r starting from 1, CA scans the ring perimeter by examining the two rows at $y \pm r$ and the column at $x + r$. If no SC is found at this stage, the ring grows by one. Otherwise, when an undiscovered SC is encountered, its index is recorded in the order vector, the cell is marked as discovered, the radius is reset to $r = 1$, and the newly found SC becomes the next reference. Since the embedding enforces nondecreasing x, any undiscovered SC must lie in the same or a rightward column. Exploiting this, the implementation prunes the left-hand perimeter entirely, reducing scan work without altering the recovered sequence. In fig. 9, the resulting magnification can be seen with recursive concentric rings indicating the growing ring and blue dots representing the strips that are taken and placed according to the order-preserving layout. For visualization only, $non-SC$ cells touched by the wavefront are labelled cyclically 1-9.

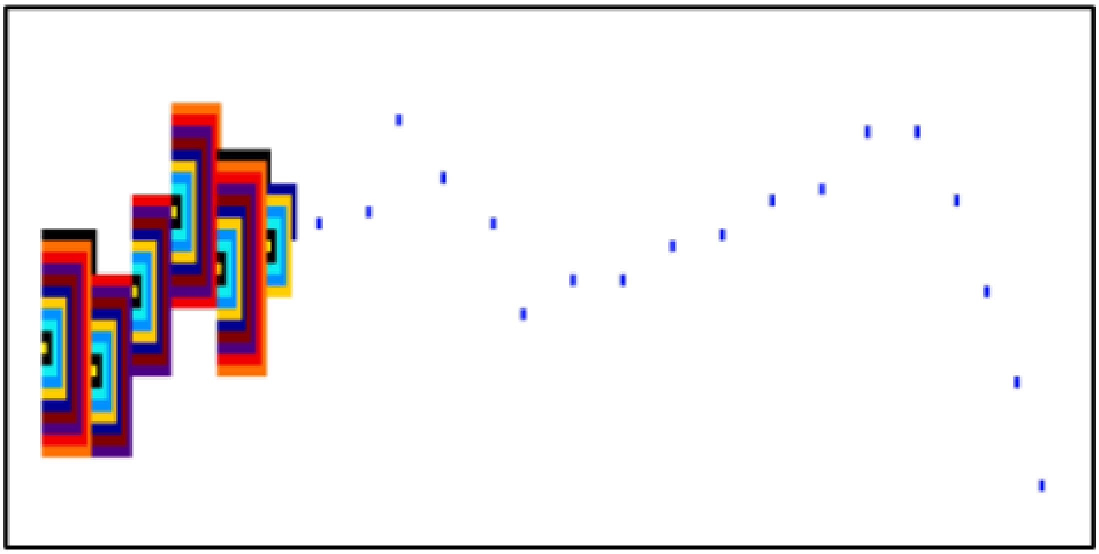


Fig. 9 *CA wavefront expansion*

The process terminates once $n - 2$ interior strips have been recovered. At that point, the only missing index in $\{1, \ldots, n - 1\}$ is inserted in position $n - 1$, and the known rightmost strip n is appended to yield the complete left-to-right order. In practice, for a fixed dissimilarity matrix D and a fixed embedding seed, the CA traversal is fully reproducible; when failures occur, they arise from weaknesses in the boundary cost rather than instability of the CA dynamics.

4.5 Deterministic Reconstruction

After the strip orders are obtained from the previous phase, the reconstruction process is determined by a deterministic path. The leftmost strip is placed at position 1. The rest of the strips are then filled sequentially, starting with position 2 to n-1 according to the order produced by CA. Then the rightmost strip is ultimately positioned at position n, hence completing the page reconstruction process.

Since the final reconstruction is entirely driven by the precomputed ordering from the previous step, the reconstruction stage itself does not introduce any visual artifacts. Consequently, any visible discontinuity in the reconstructed page indicates an incorrect ordering rather than an issue arising from the reconstruction stage. The reconstruction stage has a linear time complexity and provides insignificant computation costs to the pipeline itself. Based on this, the accuracy of the ultimate reconstruction is largely determined by the accuracy of the ordering produced by the CA, and not the reconstruction process itself.

5. Results and Discussion

We test the entire pipeline on 120 pages that have been virtually ripped into strips of size {20, 25, 30}, giving a total of 360 test cases. The pages are stratified into three content density categories to investigate difficulty: Category 1 has text-dense pages with limited large white space; Category 2 has medium density pages with vertical white lines, which frequently run down the right page; Category 3 has sparse pages with white space in the center. With this stratification, we can determine the extent to which edge information, distribution of whitespace, and page layout affect boundary matching and ordering. The approach achieves a performance of 70.74% page-level correctness across all 360 instances with single-CPU runtimes of 30-55 seconds. Accuracy is strongly tied to content density: Category 1 = 99%, Category 2 ≈ 73%, and Category 3 ≈ 40.23% exact reconstructions. Breaking the results down by content density makes the trend unambiguous: page structure rather than the number of strips governs the difficulty of reconstruction. Under the condition that the edge information has enough distinctiveness and less white space, the boundary cost function $D(i, j)$ can significantly classify the adjacent strips and thus help the CA to identify the correct order. On the other hand, the larger the whitespace and the closer the edge patterns to homogeneity, the smaller the discriminative power of the boundary cost function, thus making the reconstruction process more difficult.

On Category 1 pages, the exact reconstruction proportion has saturated and is 100% at the sample size of n=20, 100% at n=25, and decreases a little to 97% at n=30. The high foreground density supplies distinctive left-to-right boundaries, so the dissimilarity is highly discriminative, and the traversal proceeds almost in a stable left-to-right manner. Moving to Category 2, accuracy remains strong but below saturation: 75.9% at n=20, 77.6% at n=25, and 65.5% at n=30. Errors concentrate in the final segment of the page: the pipeline typically reconstructs most of the content correctly and then makes one or two swaps near the right edge, precisely where long white runs weaken the boundary signal and result in ambiguous strip alignments. By contrast, Category 3 is the most challenging: 55.2% at n=20, 34.5% at n=25, and 31% at n=30. Large uniform regions render consecutive boundaries similar, so small perturbations in the cost can flip local choices; once early placements go wrong, the resulting misorder can cascade across the page.

The results of the performance trends, as illustrated in Table 1, indicate the behavior of the proposed method among the various layout characteristics. Instead of using external baselines, such analysis is done in intra-method consistency across content categories. The outcome reveals that the technique is very stable on pages with a lot of text, moderately so on mixed format, and highly sensitive on whitespace-intensive situations. This development brings out the vastness as well as the obvious weakness of the method, as the page structure gets blurred.

Table 1 *Accuracy by category and strip count*

Category	n = 20	n = 25	n = 30	Overall
Category 1	100%	100%	97%	99%
Category 2	75.9%	77.6%	65.5%	73.0%
Category 3	55.2%	34.5%	31.0%	40.2%

We evaluate at the page level: a page is counted as correct only if every strip is in its exact position. This all-or-nothing criterion means a single misplaced strip renders the whole page incorrect; partially correct layouts, even when most of the page is right, are deliberately scored as failures. Under this metric, the CA paired with the order-preserving embedding is sound, deterministic, and efficient. The observed errors arise from the boundary dissimilarity rather than CA dynamics. Pages with long white runs or sparse content provide weak edge cues, so adjacent strips become hard to distinguish, and the cost may select the wrong neighbor. The clearest path to improvement is therefore the cost function itself. With a more informative dissimilarity, the same CA pipeline should yield consistently higher page-level correctness.

6. Conclusion

This paper demonstrates that strip-cut document reconstruction is formulated as a directed path-TSP on strip adjacencies and is solvable effectively by CA. The algorithm presented has five key steps, namely dataset construction, computation of boundary dissimilarities, order-preserving 2D embedding, CA-based traversal to reconstruct the left-to-right sequence, and deterministic page reconstruction.

Where experiments are carried out on a corpus of 120 pages, corresponding to 360 instances, the page-level performance is strong. Reconstruction accuracy seems to be strongly associated with the content density: the most text-rich pages achieve near-perfect results (99% percent in Category 1), mixed-content pages achieve strong results with moderate degradation (around 73% in Category 2), and whitespace-dominant pages are the most challenging ones (around 40.23% in Category 3).

Analysis of errors points out that failures are mostly due to ambiguities in the measure of the boundary dissimilarity when large white areas reduce the edge distinctiveness. Under these circumstances, neighboring strips are hard to distinguish, and the CA follows the cost provided, even though it is not the best one. This observation confirms that the main weakness lies in the boundary cost formulation and not in the CA mechanism itself.

Based on this insight, future developments will thus focus on the improvement of the boundary cost formulation in order to better suit the whitespace-heavy pages while maintaining the current CA-based solver. In addition to this, we will test the methodology to cross-cut documents and that of real scanned pages with noise, skew, and missing or overlapping strips.

Acknowledgement

The authors would like to thank Al-Ahliyya Amman University for its support.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Abdallah S. Hyassat, Abdel Latif Abu Dalhoum;* ***data collection:*** *Abdallah S. Hyassat, Insaf Kraidia, Iyas Qaddara;* ***analysis and interpretation of results:*** *Abdallah S. Hyassat, Abdel Latif Abu Dalhoum, Mosleh M. Abualhaj;* ***draft manuscript preparation:*** *Abdallah S. Hyassat, Abdel Latif Abu Dalhoum, Mosleh M. Abualhaj. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Biswas, A., Bhowmick, P., & Bhattacharya, B. B. (2005). Reconstruction of torn documents using contour maps. IEEE International Conference on Image Processing 2005, III–517. https://doi.org/10.1109/ICIP.2005.1530442

[2] Patel, B., & Amin, J. (2015). Reconstruction of shredded document using image mosaicing technique–A survey. International Journal of Science and Research (IJSR), 4(12), 737–740. https://doi.org/10.21275/NOV152063

[3] Chotikunnan, P., Chotikunnan, R., Puttasakul, T., Khotakham, W., Imura, P., Prinyakupt, J., et al. (2025). Noise-reduced 3D organ modeling from CT images using median filtering for anatomical preservation in medical 3D printing. Journal of Robotics and Control (JRC), 6(4), 1600–1611. https://doi.org/10.18196/jrc.v6i4.26665

[4] Kriouich, M., Sarir, H., & Louah, S. (2024). Research trends and knowledge taxonomy of artificial intelligence applications in supply chain management, logistics, and transportation: A systematic literature review and bibliometric analysis. Journal of Robotics and Control (JRC), 5(5), 1349–1364. https://doi.org/10.18196/jrc.v5i5.21859

[5] Hiari, M. O., & Shambour, Q. Y. (2025). Enhancing medication recommendations through multi-criteria collaborative filtering. Engineering, Technology & Applied Science Research, 15(5), 27181–27187. https://doi.org/10.48084/etasr.12239

[6] Abualhaj, M. M., Al-Khatib, S. N., Alsharaiah, M. A., & Hiari, M. O. (2024). Performance comparison of whale and Harris hawks optimizers with network intrusion prevention systems. Journal of Applied Data Sciences, 5(4), 1530–1538. https://doi.org/10.47738/jads.v5i4.323

[7] Abualhaj, M. M., Al-Khatib, S., Al Shafi, N., Qaddara, I., & Hyassat, A. (2025). Utilizing Gray Wolf optimization algorithm in malware forensic investigation. Journal of Computational and Cognitive Engineering. https://doi.org/10.47852/bonviewJCCE52025053

[8] Alraba'nah, Y., Al-Sharaeh, S., & Al Hindi, G. (2025). Enhancing intrusion detection using hybrid long short-term memory and XGBoost. Journal of Soft Computing and Data Mining, 6(1), 247–261. https://doi.org/10.30880/jscdm.2025.06.01.016

[9] Manzoni, L., Mariot, L., & Menara, G. (2026). Combinatorial designs and cellular automata: A survey. Discrete Applied Mathematics, 379, 656–674. https://doi.org/10.1016/j.dam.2025.10.014

[10] Rollier, M., Zielinski, K. M. C., Daly, A. J., Bruno, O. M., & Baetens, J. M. (2025). A comprehensive taxonomy of cellular automata. Communications in Nonlinear Science and Numerical Simulation, 140, 108362. https://doi.org/10.1016/j.cnsns.2024.108362.

[11] Alabdullah, B., Alsolai, H., Alhayan, F., Ikram, A., Almjally, A., Shehab, M., & Albahar, M. A. (2025). A novel image encryption framework using Wireworld cellular automaton and hybrid chaotic maps for enhanced security. PLoS One, 20(11), e0332480. https://doi.org/10.1371/journal.pone.0332480

[12] Mariot, L. (2023). Enumeration of maximal cycles generated by orthogonal cellular automata. Natural Computing, 22(3), 477–491. https://doi.org/10.1007/s11047-022-09930-1

[13] Wali, A. (2025). CA-NN: A cellular automata neural network for handwritten pattern recognition. Natural Computing, 24(2), 173–180. https://doi.org/10.1007/s11047-022-09937-8

[14] Marzoug, R., Lakouari, N., Pérez Cruz, J. R., Castillo-Téllez, B., Mejía-Pérez, G. A., & Bamaarouf, O. (2025). Cellular automata for optimization of traffic emission and flow dynamics in two-route systems using feedback information. Infrastructures (Basel), 10(5). https://doi.org/10.3390/infrastructures10050120

[15] Earle, S., & Togelius, J. (2025). Autoverse: Evolving symbolic neural cellular automata environments to train player agents. Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '25 Companion), 175–178. https://doi.org/10.1145/3712255.3726722

[16] Fonseca i Casas, P. (2025). The multi-n-dimensional cellular automaton: A unified framework for tensorial, discrete, and continuous simulations—A computational definition of time. Complexity, 2025(1), 3088010. https://doi.org/10.1155/cplx/3088010

[17] Kanada, Y. (1994). Asynchronous 1d cellular automata and the effects of fluctuation and randomness. Proceedings of the Fourth Conference on Artificial Life (A-Life IV). MIT Press.

[18] Castillo-Ramirez, A., & Magaña-Chavez, M. G. (2025). A study on the composition of elementary cellular automata. In A. Adamatzky, G. Ch. Sirakoulis, & G. J. Martinez (Eds.), Advances in cellular automata: Volume 1: Theory (pp. 347–373). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78757-7_12

[19] George, S. N., Augustine, N., & Pattathil, D. P. (2015). Audio security through compressive sampling and cellular automata. Multimedia Tools and Applications, 74(23), 10393–10417. https://doi.org/10.1007/s11042-014-2172-2

[20] Hong, Z., & Bian, F. (2008). Novel ant colony optimization for solving traveling salesman problem in congested transportation system. 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 122–125. https://doi.org/10.1109/PACIIA.2008.259

[21] Bacha, A. M., Zamoum, R. B., & Lachekhab, F. (2025). Machine learning paradigms for UAV path planning: Review and challenges. Journal of Robotics and Control (JRC), 6(1), 215–233. https://doi.org/10.18196/jrc.v6i1.24097

[22] Bock, S., Bomsdorf, S., Boysen, N., & Schneider, M. (2025). A survey on the traveling salesman problem and its variants in a warehousing context. European Journal of Operational Research, 322(1), 1–14. https://doi.org/10.1016/j.ejor.2024.04.014

[23] Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., & Belongie, S. (2007). Generalized non-metric multidimensional scaling. Artificial Intelligence and Statistics, 11–18.

[24] Justino, E., Oliveira, L. S., & Freitas, C. (2006). Reconstructing shredded documents through feature matching. Forensic Science International, 160(2), 140–147. https://doi.org/10.1016/j.forsciint.2005.09.001

[25] Richter, F., Ries, C. X., Cebron, N., & Lienhart, R. (2013). Learning to reassemble shredded documents. IEEE Transactions on Multimedia, 15(3), 582–593. https://doi.org/10.1109/TMM.2012.2235415

[26] Ukovich, A., Ramponi, G., Doulaverakis, H., Kompatsiaris, Y., & Strintzis, M. G. (2004). Shredded document reconstruction using MPEG-7 standard descriptors. Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 334–337. https://doi.org/10.1109/ISSPIT.2004.1433788

[27] Lin, H.-Y., & Fan-Chiang, W.-C. (2012). Reconstruction of shredded document based on image feature matching. Expert Systems with Applications, 39(3), 3324–3332. https://doi.org/10.1016/j.eswa.2011.09.019

[28] Prandtstetter, M., & Raidl, G. R. (2008). Combining forces to reconstruct strip shredded text documents. In M. J. Blesa, C. Blum, C. Cotta, A. J. Fernández, J. E. Gallardo, A. Roli, & M. Sampels (Eds.), Hybrid metaheuristics (pp. 175–189). Springer Berlin Heidelberg.

[29] Marques, M. A. O., & Freitas, C. O. A. (2009). Reconstructing strip-shredded documents using color as feature matching. Proceedings of the 2009 ACM Symposium on Applied Computing, 893–894. https://doi.org/10.1145/1529282.1529475

[30] Shihab, R. M., & Thorig, I. (2024). Gauging the effectiveness of strip cut un-shredding techniques on Dhivehi text-based documents. AIP Conference Proceedings, 3245(1), 050006. https://doi.org/10.1063/5.0231940

[31] Sleit, A., Massad, Y., & Musaddaq, M. (2013). An alternative clustering approach for reconstructing cross cut shredded text documents. Telecommunication Systems, 52(3), 1491–1501. https://doi.org/10.1007/s11235-011-9626-x

[32] Prandtstetter, M., & Raidl, G. R. (2009). Meta-heuristics for reconstructing cross cut shredded text documents. Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation, 349–356. https://doi.org/10.1145/1569901.1569950

[33] Xu, H., Zheng, J., Zhuang, Z., & Fan, S. (2014). A solution to reconstruct cross-cut shredded text documents based on character recognition and genetic algorithm. Abstract and Applied Analysis, 2014, 829602. https://doi.org/10.1155/2014/829602

[34] Liu, H., Cao, S., & Yan, S. (2011). Automated assembly of shredded pieces from multiple photos. IEEE Transactions on Multimedia, 13(5), 1154–1162. https://doi.org/10.1109/TMM.2011.2160845

[35] Gong, Y. J., Ge, Y. F., Li, J. J., Zhang, J., & Ip, W. H. (2016). A splicing-driven memetic algorithm for reconstructing cross-cut shredded text documents. Applied Soft Computing, 45, 163–172. https://doi.org/10.1016/j.asoc.2016.03.024

[36] Chen, J., Ke, D., Wang, Z., & Liu, Y. (2018). A high splicing accuracy solution to reconstruction of cross-cut shredded text document problem. Multimedia Tools and Applications, 77(15), 19281–19300. https://doi.org/10.1007/s11042-017-5389-z

[37] Xu, H., Zheng, J., Zhuang, Z., & Fan, S. (2014). A solution to reconstruct cross-cut shredded text documents based on character recognition and genetic algorithm. Abstract and Applied Analysis, 2014(1), 829602. https://doi.org/10.1155/2014/829602

[38] Zhang, Z., Zou, J., Yang, S., Zheng, J., Gong, D., & Pei, T. (2022). A reconstruction method for cross-cut shredded documents based on the extreme learning machine algorithm. Soft Computing, 26(22), 12851–12862. https://doi.org/10.1007/s00500-022-07311-5