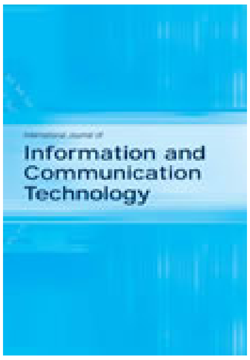


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Adaptive genetic algorithm for multi-topic polyphonic music generation

Huiyan Zeng

DOI: 10.1504/IJICT.2025.10074814

Article History:

Received:	30 August 2025
Last revised:	02 October 2025
Accepted:	02 October 2025
Published online:	12 December 2025

Adaptive genetic algorithm for multi-topic polyphonic music generation

Huiyan Zeng

School of Tourism Home Economics and Arts,
Qingyuan Polytechnic,
Qingyuan, 511500, China
Email: zenghuiyan666@163.com

Abstract: Automatic composition struggles to balance multi-theme planning with strict polyphonic constraints. To address this, this paper proposes an adaptive genetic framework for multi-topic polyphonic music generation. In the scheme, first a bar-aligned, voice-aware representation is prepared with tonal cues and a theme schedule. Then a domain-aware evolutionary core explores the search space through voice-preserving crossover, musically constrained mutation, and lightweight local repair around exposed phrases. Finally, a composite evaluator guides selection while an adaptive controller adjusts operator rates using diversity and stagnation signals. Experiments on chorale, chamber, and modern tonal sets show fewer rule violations, higher consonance with 83% vertical consonance and tonal stability, stronger theme recognisability, and faster convergence without extra runtime. The approach delivers structured, stylistically credible music with strong controllability, clear diagnostics, and room for interactive use.

Keywords: algorithmic composition; polyphonic music; adaptive genetic algorithm; thematic scheduling.

Reference to this paper should be made as follows: Zeng, H. (2025) 'Adaptive genetic algorithm for multi-topic polyphonic music generation', *Int. J. Information and Communication Technology*, Vol. 26, No. 45, pp.83–100.

Biographical notes: Huiyan Zeng received her Master's degree from Hunan Normal University, China in 2011. Currently, she works as a Lecturer at the School of Tourism Home Economics and Arts, Qingyuan Polytechnic. Her research interests include polyphonic music, aesthetic education and thematic scheduling.

1 Introduction

Algorithmic composition has long pursued a delicate balance: granting machines enough freedom to invent new musical ideas while enforcing the structural discipline that gives polyphonic music its coherence. In monophonic settings, this tension is manageable; sequence models can unfold a line that sounds plausible, and post-filters rarely derail the result (Abbasi et al., 2023). Polyphony is different. Multiple independent voices must breathe together yet avoid collapsing into parallel motion or harmonic clutter; melodic contours should remain singable even as harmony modulates; rhythmic textures must interlock without turning mechanical. These requirements become sharper when a piece deliberately incorporates more than one thematic idea. Multi-topic writing-where distinct motives alternate, overlap, transform, or migrate across voices-demands not only local correctness (e.g., legal intervals at each instant) but also mid-range continuity (e.g., how a theme returns after modulation) and long-range narrative (e.g., a pattern of tension and release that listeners can track across sections) (Acharya and Kumar, 2021). Contemporary data-driven systems excel at capturing local correlations in symbolic corpora, yet they typically lack an explicit handle on hard musical constraints or on the choreography of themes over time. Conversely, rule-based engines preserve contrapuntal hygiene but tend to produce brittle or predictable textures, especially when asked to juggle more than one theme or style. What is missing is a search strategy that can weave between these extremes-capable of maintaining formal guardrails in real-time, yet plastic enough to explore nontrivial reorganisations of material as themes enter, depart, and combine (Ali et al., 2018).

We address this need with an adaptive genetic algorithm designed specifically for multi-topic polyphonic generation (Hao et al., 2022). The method treats candidate scores as chromosomes built from bar-level building blocks, each block encoding synchronised voices with explicit onset grids, pitch classes, ranges, and thematic tags. Instead of using a fixed schedule for crossover and mutation, the algorithm adapts these rates to the population's evolving state: when diversity drops or best-fitness stagnates, exploration increases; when promising structures begin to coalesce, exploitation intensifies through more conservative crossover and local refinements. Selection favours individuals that optimise a composite fitness with four intertwined aspects. First, contrapuntal legality captures voice-leading quality, penalising voice crossing, hidden parallels, bad accented dissonances, and unidiomatic leaps without forcing a single school of species counterpoint (Hong et al., 2018); this allows the system to respect stylistic breadth while discouraging canonical errors. Second, harmonic and tonal stability measures how well vertical sonorities arise at musically sensible metrical points and how modulations proceed along plausible pathways rather than random chromatic drift. Third, theme and style consistency gauges whether required topics actually occur where scheduled, whether they undergo accepted transformations (augmentation, inversion, register shift) without losing identity, and whether the overall idiom remains close to a reference corpus (Ju and Liu, 2019); this component leverages symbolic language modelling for soft stylistic proximity while leaving room for novel combinations. Fourth, diversity and developmental balance ensure that the population does not collapse into clones, and that sections exhibit enough variation to sustain listener interest without erasing the piece's thematic backbone. Beyond adaptive rates, the engine employs niche preservation to protect distinct 'species' of solutions, a light-weight memetic step to polish local voice-leading around cadence points, and domain-specific repair operators that resolve illegal intervals, restore ranges, or realign chord tones on tactically important beats after recombination. A topic scheduler coordinates when each theme enters and how overlapping themes are spliced across voices; it can hard-pin critical entrances (e.g., a subject at bar 1 and 9) while allowing the search to determine secondary placements and countersubject pairings. This combination of adaptive exploration, soft style guidance, and explicit musical repair yields a search landscape where large-scale structure can emerge without sacrificing the grammar of polyphony (Kaliakatsos-Papakostas et al., 2020). Relative to prior genetic music systems, the design adds an adaptive controller driven by diversity and stagnation signals, a policy scheduler that hard pins or softly windows thematic entrances, and a lightweight local repair around exposed cadential regions. These choices reduce infractions, preserve recognisability, and sustain long-range form.

The remainder of the paper situates this approach within recent work and substantiates its effectiveness with quantitative and qualitative evidence. We curate symbolic corpora that span contrasting idioms-contrapuntal chorales, small-ensemble chamber textures, and modern tonal miniatures-and segment them into thematic units to train the scheduler and provide stylistic references (Kuptametee et al., 2023). Experimental design reflects both the algorithmic and artistic nature of the problem. Objective analyses report rates of contrapuntal infractions per hundred beats, tonal drift measures based on key profiles, rhythmic entropy and inter-voice independence indices, and explicit coverage statistics for the requested themes across sections; convergence behaviour is tracked to compare the adaptive strategy with a non-adaptive GA and with alternative heuristics such as particle swarm and simulated annealing (Luo and Oyedele, 2021). Human-in-the-loop evaluation complements these numbers: in a double-blind listening study, participants rate perceived coherence, thematic recognisability, and overall musicality while also indicating whether they can detect the presence and interplay of multiple topics (Luo et al., 2020). Ablations isolate the value of each ingredient-adaptive rate control, repair operators, niche maintenance, and thematic scheduling-so that gains cannot be attributed merely to more compute or to corpus idiosyncrasies (Ma and He, 2019). Our contributions are therefore threefold: an adaptive genetic framework that explicitly couples thematic planning with polyphonic constraints, a set of musically grounded objectives and repairs that turn brittle rule-checking into constructive guidance, and an evaluation protocol that acknowledges both the craft of counterpoint and the listener's experience (Ma et al., 2025). Section 2 reviews the relevant background in symbolic representations, polyphonic rule systems, and evolutionary search for creative tasks. Section 3 details the proposed method, from chromosome layout and adaptive control signals to fitness aggregation and repair logic. Section 4 presents results, ablations, convergence profiles, and listening study outcomes, with musical excerpts to illustrate how themes transit and combine. Section 5 concludes by reflecting on limitations-such as the handling of extreme syncopation or modal mixtures-and on extensions toward interactive co-creation and real-time accompaniment (Mokshin et al., 2019).

2 Relevant technologies

2.1 Symbolic representations and pre-processing

Multi-topic polyphonic generation is grounded in a symbolic pipeline that preserves timing precision, voice identity, pitch spelling, and thematic labels. Common encodings (e.g., MIDI, MusicXML, piano-roll, event streams) differ in granularity, but a consistent abstract layer helps subsequent optimisation. Two representation choices matter most for an evolutionary search:

1 an onset grid that aligns events to metrical anchors while tolerating expressive deviation

2 a pitch representation that permits fast tonal computations while remaining invertible to staff notation (Ness and Dysthe, 2020).

A convenient pitch abstraction is the pitch-class map that normalises enharmonic choices and facilitates consonance checks:

$$pc(n) = n \bmod 12 \tag{1}$$

where n denotes the MIDI note number or any integer semitone index, and $pc(n) \in \{0, \ldots, 11\}$ is the pitch class used by interval and consonance logic (Prado et al., 2020).

Before optimisation, expressive onsets are mapped to a bar-level grid $\{g_k\}$ (e.g., 1/12 or 1/16 of a bar) to enable bar-wise recombination (Que et al., 2018). Quantisation is framed as a least-squares fit with robust tie-breaking for sustained tones:

$$Q = \frac{1}{T}\sum_{t=1}^{T} \min_k \left\| o_t - g_k \right\|^2 \tag{2}$$

where o_t is the original onset time of the t^{th} event, g_k is the k^{th} grid point in the current bar, T is the number of events under consideration, and Q is the average quantisation error minimised during pre-processing.

Tonality cues are obtained by correlating pitch-class histograms with a rotating key profile (e.g., Krumhansl-Schmuckler style). Let $h \in \mathbb{R}^{12}$ be the normalised pitch-class histogram over a sliding window and $c \in \mathbb{R}^{12}$ a major or minor template. Rotations $R_r c$ enumerate all keys:

$$k(r) = \frac{h^{\top}\left(R_r c\right)}{\|h\|_2 \left\|R_r c\right\|_2}, \quad r \in \{0, \ldots, 11\} \tag{3}$$

where $k(r)$ is the cosine correlation for rotation r, R_r circularly shifts the template by r pitch classes, and the maximiser $\operatorname{argmax}_r k(r)$ identifies the most likely tonic centre used later by harmonic evaluators and topic scheduling. One bar comprises four voices aligned on a twelfth grid. Voice S enters on beat one with a stepwise ascent outlining tonic. Voice A sustains a third above. Voice T moves in contrary motion toward the cadence. Voice B marks roots on strong ticks. Theme A tag active.

Beyond these basics, pre-processing also annotates metrical strength (downbeat, secondary accents), voice ranges, instrument registers, and theme IDs. These tags allow bar-wise building blocks to remain musically meaningful under crossover (Resende and Drummond, 2018).

2.2 *Formalisation of polyphony and contrapuntal constraints*

Polyphony imposes local and mid-range relationships among voices. Rather than hard-rejecting candidates, constraint families are converted into differentiable or piecewise-linear penalties aggregated into a contrapuntal energy (Xue et al., 2021). Let $\mathcal{V}$ index voices and t traverse synchronised frames on the onset grid. A compact energy reads:

$$E_{ctr} = \sum_{t} \sum_{(u,v)\in\mathcal{P}} \left(\lambda_1 \Pi_{5,8}^{(u,v,t)} + \lambda_2 \tilde{\Pi}_{5,8}^{(u,v,t)} + \lambda_3 X^{(u,v,t)} \right) + \sum_{v\in\mathcal{V}} \sum_{t} \left(\lambda_4 L^{(v,t)} + \lambda_5 D^{(v,t)} \right) \tag{4}$$

where

1 $\mathcal{P}$ is the set of unordered voice pairs

2 $\Pi_{5,8}^{(u,v,t)} \in \{0,1\}$ flags exact parallel fifths/octaves between voices u, v at frame t and $t-1$

3 $\tilde{\Pi}_{5,8}^{(u,v,t)} \in \{0,1\}$ flags hidden (direct) fifths/octaves caused by similar motion into perfect consonances

4 $X^{(u,v,t)} \in \{0, 1\}$ flags voice crossing between u and v

5 $L^{(v,t)} \geq 0$ penalises unidiomatic leaps in a single voice (e.g., leaps exceeding a sixth without compensating contrary step);

6 $D^{(v,t)} \geq 0$ penalises accented dissonances not prepared/resolved by step;

7 $\lambda_1, \ldots, \lambda_5 > 0$ weight penalties by stylistic tolerance (e.g., stricter in species-like textures, looser in modern tonal miniatures).

Voice-leading smoothness is treated separately to encourage cantabile lines. A convex power cost balances tolerance for small steps with discouragement of jagged contours:

$$E_{sm} = \sum_{v \in \mathcal{V}} \sum_{t} \left| p_{v,t} - p_{v,t-1} \right|^{\alpha}, \quad \alpha \geq 1 \tag{5}$$

where $p_{v,t}$ is the MIDI pitch (or pitch class with register) of voice v at frame t, and α controls the step/skip trade-off (e.g., α = 1 for L_1 smoothness, α = 2 for quadratic smoothing).

Vertical sonorities are assessed by consonance vectors derived from interval-class statistics. Let S_t be the set of sounding pitches (with multiplicity) at frame t, and $icv(S_t) \in \mathbb{R}^6$ its interval-class vector. With style weights $w \in \mathbb{R}^6$ (favouring 3rds/6ths, penalising tritones), a harmonic consonance score per frame is

$$C_t = w^{\top} icv(S_t), \text{ and } H_{vert} = \frac{1}{T} \sum_{t=1}^{T} C_t \tag{6}$$

where C_t measures the consonance quality of the simultaneity at t, and H_{vert} is the average vertical consonance over the passage.

Tonal stability requires metrical sensitivity: chord tones on strong beats should align with the current key trajectory. Using the key correlation $k(r)$ from §2.1 to determine the active key κ_t, accented instants $\mathcal{A} \subset \{1, \ldots, T\}$ are emphasised:

$$H_{tonal} = \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} 1\left[pc(S_t) \subseteq \mathcal{T}(\kappa_t) \right] \tag{7}$$

where $\mathbf{1}[\cdot]$ is the indicator of chord-tone membership, $pc(S_t)$ the multi-set of pitch classes at t, and $\mathcal{T}(\kappa_t)$ the admissible chord-tone set under key κ_t (including diatonic alterations per style).

The penalty terms E_{ctr} and E_{sm} and the harmony scores H_{vert}, H_{tonal} are normalised before entering multi-objective fitness. Explicit repair operators-interval fix-ups, range restoration, metrical realignment-act after recombination to reduce E_{ctr} without erasing emergent structure (Yadav et al., 2022).

2.3 Evolutionary search and adaptive genetic control

Genetic algorithms become effective for symbolic composition when the chromosome layout respects bar boundaries and voice identity (Yuan, 2024). Individuals are built from bar-level blocks with synchronised voices and topic tags. Fitness aggregates contrapuntal, harmonic, thematic, and diversity components. A practical scalarisation for selection is:

$$F = w_1 \left(1 - \hat{E}_{ctr}\right) + w_2 \hat{H}_{vert} + w_3 \hat{H}_{tonal} + w_4 S_{topic} + w_5 D_{pop} \tag{8}$$

where hats denote min-max normalised terms; S_{topic} is the theme/style consistency score. D_{pop} is the population diversity statistic below; and $w_1, \ldots, w_5 \geq 0$ are user-set or auto-tuned weights (later experiments vary them in ablations).

Population diversity is estimated by an average pairwise distance that mixes pitch, rhythm, and harmonic differences. Let $d(.,.)$ be a composite distance (e.g., weighted edit distance on event streams plus chord-set Jaccard), and $\{x_i\}_{i=1}^{N}$ the population:

$$D_{pop} = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} d\left(x_i, x_j\right) \tag{9}$$

where N is the population size and D_{pop} rises with structural variety, discouraging premature convergence.

Adaptive control links operator rates to diversity and stagnation statistics. Let g be the generation index, $D^{(g)}$ the measured diversity, and D^* a target range (e.g., the running median of early generations). The crossover probability is raised when diversity collapses and tempered when diversity exceeds the target:

$$p_c^{(g+1)} = clip\left(p_c^{(g)} \exp\left(\alpha_c \frac{D^\star - D^{(g)}}{D^\star + \grave{o}} \right), p_c^{\min}, p_c^{\max} \right) \tag{10}$$

where $p_c^{(g)}$ is the current crossover rate, $\alpha_c > 0$ a sensitivity constant, $\epsilon > 0$ a small stabiliser, and $clip(x, a, b) = \min\{b, \max\{a, x\}\}$ enforces bounds $[p_c^{\min}, p_c^{\max}]$.

Mutation responds to fitness stagnation and low diversity. Let $s^{(g)} \in [0, 1]$ measure stagnation (e.g., a sigmoid of generations since last improvement in the elite), then

$$p_m^{(g+1)} = clip\left(p_m^{(g)} \exp\left(\alpha_m \frac{D^\star - D^{(g)}}{D^\star + \grave{o}} \right)\left(1 + \beta_m s^{(g)}\right), p_m^{\min}, p_m^{\max} \right) \tag{11}$$

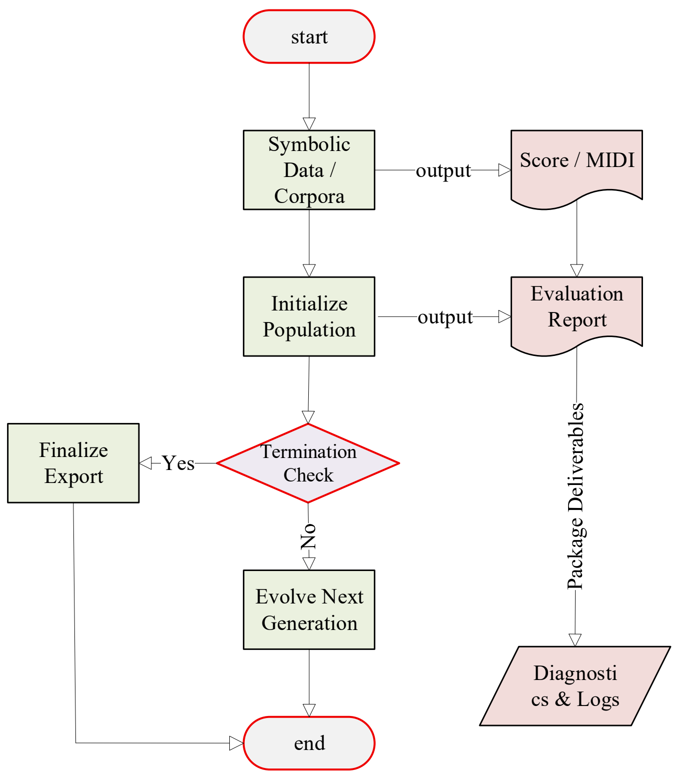
where $p_m^{(g)}$ is the mutation rate at generation g, α_m, $\beta_m > 0$control diversity- and stagnation-driven adjustments, and the same *clip* keeps mutation within $[p_m^{\min}, p_m^{\max}]$.

To protect alternative structural 'species' (e.g., different ways of staging theme B against theme A), niche sharing rescales fitness by local density:

$$\hat{f}_i = \frac{f_i}{\sum_{j=1}^{N} sh\left(d\left(x_i, x_j\right)\right)} \quad sh(d) = \begin{cases} 1 - \left(\frac{d}{\sigma_{share}}\right)^{\gamma}, & d < \sigma_{share}, \\ 0, & d \geq \sigma_{share}. \end{cases} \tag{12}$$

where f_i is the raw scalar fitness of individual i, $d(.,.)$ is the same structural distance used in D_{pop}, $\sigma_{share} > 0$ is the niche radius, $\gamma \geq 1$ shapes the sharing kernel, and $\hat{f}_i$ is the density-adjusted value used by selection. Primary crossover swaps bar aligned blocks while preserving voice identity and meter. Mutation employs rhythm splits or merges on weak positions, pitch substitutions within the active key neighbourhood, and idiomatic ornaments. Typical probabilities favour crossover near one half and mutation near one fifth. Local repair triggers after every recombination.

Figure 1 System architecture for adaptive genetic multi-topic polyphonic generation (see online version for colours)



Operator semantics are domain-aware. Crossover swaps bar-aligned blocks but preserves voice indices; topic and meter labels are inherited to maintain alignment with the scheduler (Zhang and Yu, 2021). Mutation covers rhythm micro-edits (duration splits/merges), pitch-set substitutions drawn from the current key neighbourhood, and theme-specific ornamentation patterns. A memetic local step (two to four frames around cadential points) runs a short greedy fix for illegal intervals and off-accent dissonances;

its cost is lower than a full repair pass and helps decrease E_{ctr} without disrupting global form. As summarised in Figure 1, the box labelled evolve next generation expands to the sequence selection, crossover, mutation, repair/memetic, optional adaptive update and candidates are then evaluated before the termination check. When the check returns yes, the pipeline proceeds to finalise and export (emitting the score/MIDI and the evaluation report), whereas no returns control to the evolutionary loop. The right-hand branch (diagnostics and logs) collects convergence curves and rule-violation reports that inform later analysis and tuning (Zhang et al., 2024).

2.4 Thematic modelling, style metrics, and schedule alignment

Multi-topic writing hinges on two complementary notions: recognisability of each theme when it appears, and coherent transitions when themes overlap or exchange roles across voices. Symbolic language modelling contributes soft stylistic pressure without over-constraining the search, while explicit alignment measures verify whether topic commitments are met (Zhou et al., 2021).

Given a normalised embedding $\phi(\cdot)$ of melodic fragments (e.g., learned from n-gram or transformer-style encoders but frozen at inference), the theme-match score between a stored prototype Θ and a candidate segment $x_{\pi:\pi+L}$ uses the best-window cosine:

$$S_{theme} = \max_{\tau\in\mathcal{W}} \frac{\phi(\Theta)^{\top}\phi\left(x_{\tau:\tau+L}\right)}{\left\|\phi(\Theta)\right\|_2 \left\|\phi\left(x_{\tau:\tau+L}\right)\right\|_2}. \tag{13}$$

where $\mathcal{W}$ enumerates window starts within the target section, L is the theme length in frames or events, and $S_{theme} \in [-1, 1]$ rises with recognisability under permissible transformations (transposition and register shifts are handled by pitch-class or transposition-invariant embeddings). Example schedule pins theme A at bars one and nine in soprano and alto while opening a soft window for theme B between bars five and eight. When multiple entrances qualify in the same region, priority favours pinned bars then higher recognisability then metrical alignment, breaking remaining ties by diversity.

Schedule adherence compares desired topic entrances with their realised counterparts. Let the plan be a set of tuples $\Pi = \{(topic\ \ell, bar\ b, voice\ set\ \mathcal{V}_\ell)\}$. Realised entrances $\hat{\Pi}$ are extracted by a detector that thresholds S_{theme} and checks voice tags. A weighted deviation cost is

$$E_{sched} = \frac{1}{|\Pi|} \sum_{(\ell,b,\mathcal{V}_\ell)\in\Pi} \min_{(\ell,\hat{b},\hat{\mathcal{V}}_\ell)\in\hat{\Pi}} \left(\eta_b\left|b-\hat{b}\right| + \eta_v\, sym0diff\left(\mathcal{V}_\ell, \hat{\mathcal{V}}_\ell\right)\right). \tag{14}$$

where η_b, η_v > 0 balance timing vs. voicing accuracy, and *sym*0*diff* is the symmetric-difference cardinality between planned and realised voice sets.

Style proximity is kept soft to avoid pastiche. A perplexity-like pseudo-likelihood over event sequences (e.g., pitch-class and duration tokens) provides a scalar guide:

$$PPL(x) = \exp\left(-\frac{1}{M}\sum_{t=1}^{M}\log p\left(x_t \middle| x_{<t}\right)\right). \tag{15}$$

where x is the tokenised sequence of length M, $p(\cdot)$ is the frozen style model's conditional probability, and lower *PPL* indicates closer adherence to the stylistic corpus.

Rhythmic independence across voices is desirable for texture clarity. Let r_v be a binary activation vector for voice v over the grid (1 if a note onset occurs). Inter-voice independence is summarised via average pairwise correlation:

$$I_{rh} = 1 - \frac{2}{|\mathcal{V}|(|\mathcal{V}|-1)} \sum_{1 \le u < v \le |\mathcal{V}|} \frac{(r_u - \overline{r_u})^\top (r_v - \overline{r_v})}{\|r_u - \overline{r_u}\|_2 \|r_v - \overline{r_v}\|_2}. \tag{16}$$

where $\overline{r_v}$ is the mean of r_v, the fraction 1 – · turns correlation into independence higher is better, and the sum averages all voice pairs.

Tonal drift is quantified by divergence between instantaneous pitch-class distributions and the key template selected. Using softmax-normalised histograms h_t and the active template $R_{\kappa_t} c$, a Kullback-Leibler-style measure is:

$$\Delta_{tonal} = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{12} h_t(i) \log \frac{h_t(i)}{[R_{\kappa_t} c](i) + \varepsilon}. \tag{17}$$

where $h_t(i)$ is the probability mass of pitch class i at time t, $\varepsilon > 0$ prevents division by zero, and lower Δ_{tonal} indicates better tonal anchoring.

The thematic and style terms above contribute to the scalar fitness F in §2.3 or to separate objectives in a multi-objective variant. In both cases, the scheduler's hard pins (e.g., 'theme A at bars 1 and 9') are respected by constraining mutation around those bars to ornamentation-only operations, while crossover is disallowed to split a pinned bar across parents.

3 Adaptive genetic framework for multi-topic polyphonic generation

3.1 Chromosome design and initialisation

The framework treats a score as a sequence of bar-aligned building blocks in which all voices are synchronised on a fixed onset grid. Each block carries both musical events and metadata: meter position, active key hypothesis, local harmonic function, voice identifiers and register limits, metrical accents, and topic tags that indicate whether a theme, countersubject, bridge, or neutral material is expected to occur. Voice identity is never anonymous; every event is bound to a named voice so that later operations can enforce range and independence rules. This representation deliberately avoids note-by-note chromosomes. A bar-level gene is long enough to express a harmonic gesture and a short voice-leading move while still being small enough to recombine without tearing larger sections apart.

A chromosome concatenates these blocks across the full piece length. The number of voices, their default ranges, and the target density profile are fixed by configuration. Blocks are allowed to differ in texture, but the bar grid and the mapping between voices and staff lines remain constant to keep edits predictable. Topic tags are embedded in the gene header and remain bound to the gene through crossover and mutation, which makes it possible to respect the scheduler's plan without imposing hard locks on the internal notes.

Initialisation follows a mixed strategy that balances plausibility and diversity. A portion of the first generation is seeded by sampling from a curated block library extracted from the symbolic corpus. Another portion is synthesised procedurally by combining rhythm templates, chordal skeletons consistent with the current key, and melodic contours shaped by register and interval statistics of the target idiom. The remaining portion is produced through style-constrained noise: random rhythms on the grid, pitch material drawn from the key neighbourhood, and immediate repair to remove obvious infractions. All three routes populate a diversity buffer that is screened by fast checks for range violations, duplicate blocks, and trivial repetitions across voices. The goal at this stage is not perfection but a population with varied structural hypotheses so that later generations can explore with meaningful recombinations.

Topic planning influences the seed. Bars marked as hard pins receive blocks that already express the required theme or a near variant. Windows marked as soft carry neutral material with sufficient rhythmic space to admit a later entrance. Bars surrounding cadences receive blocks with clear harmonic bearings and strong accents to simplify later memetic polishing. The initialisation step also records a per-chromosome sketch of tension over time derived from accent density and harmonic rhythm, which is later used to guide selection when two individuals are otherwise tied.

3.2 Adaptive search process

One generation advances through a disciplined cycle: selection, crossover, mutation, local repair, evaluation, and rate adaptation. Selection is rank-based with elitism to secure steady progress. Individuals are compared not only by their composite fitness but also by their distance to neighbours in the current population. The distance function considers contour similarity within each voice, onset patterns across voices, and the chordal backbone on accented beats. This simple device prevents the pool from collapsing into near duplicates and protects promising but distinct structural ideas while global fitness is still volatile.

Crossover is performed at bar boundaries and is voice-aware. When two parents exchange material, the operation swaps aligned blocks while preserving the mapping from voice identifiers to staff positions. Topic and meter labels move with the blocks so that offspring remain consistent with the scheduler. When a seam falls in the middle of a harmonic motion, a small smoothing step adjusts the entry or exit sonority to avoid abrupt dislocations. Cadential bars are treated as atomic units and are not split. This policy retains enough continuity for large-scale form to emerge while still enabling meaningful recombination of local gestures.

Mutation operates as a set of domain-specific micro-edits rather than as blind noise. Rhythmic edits split or merge durations on weak positions to create syncopation or to tighten alignment with accents. Pitch edits sample from a key neighbourhood and privilege chord tones at strong beats, while permitting controlled colour tones at weak positions. A third family of edits adds idiomatic ornaments such as passing motions, neighbour figures, and suspensions. All edits respect voice ranges and avoid direct crossings. Mutations near hard-pinned theme bars are conservative: ornamentation and register shifts are allowed, wholesale substitution is not.

Local repair follows every recombination step but remains lightweight. The repair pass scans short windows around cadences and other structurally exposed positions. In each window the pass searches for the lowest-cost correction that removes illegal intervals, restores stepwise resolution after suspensions, and re-anchors chord tones on metrically strong ticks. The routine is greedy and bounded in time so that generations remain fast, yet it removes a significant portion of the infractions that would otherwise accumulate and stall progress.

Evaluation aggregates contrapuntal legality, harmonic soundness, tonal stability on accented positions, thematic recognisability, schedule adherence, and population diversity. Each individual receives a compact diagnostic trace that lists violation counts by category, theme coverage across sections, and a coarse tension profile over time. These traces drive not only selection but also the adaptive controller that tunes exploration pressure.

Adaptation is based on two signals: diversity and stagnation. When diversity falls below a target band, the controller raises crossover probability and enables more adventurous rhythmic and pitch edits. When the best fitness has not improved for several generations, mutation intensity rises and the system widens the accepted key neighbourhood to encourage modulations and alternative voice-leading plans. When progress resumes, rates are gradually tempered to consolidate structure. The controller also applies local rules: mutation intensity is capped around hard-pinned bars and around cadence windows that have already been polished, while it is allowed to spike in transitional passages that the scheduler marks as flexible.

Termination relies on a combination of stopping rules. The run ends when the composite objective stabilises within a tight band for a sustained number of generations, when the maximum generation count is reached, or when all scheduled themes have been placed with acceptable recognisability and spacing. The final step collects the best individual, reconstructs a clean score with staff notation and dynamics inferred from accent patterns, renders audio for listening tests, and exports diagnostics for the experimental section.

3.3 Thematic control and policy layer

The policy layer governs when and where thematic material enters, how themes interact, and how structural goals shape the search. It does not dictate every note. Instead, it sets commitments that the evolutionary process must honour while leaving enough freedom for variation and development.

A schedule is prepared before the run. It lists mandatory entrances for each theme, optional echoes or answers, and windows where overlap is allowed. Every entry associates a bar index with a voice set. Some entries are hard, which means the theme must appear in that exact bar and in at least one of the intended voices. Others are soft and merely describe a preferred window. The schedule also names transitional regions in which modulation or texture change is desirable and where the controller is free to raise exploration pressure.

During initialisation, bars tagged as hard pins receive prototypes or close derivatives from the theme library. During crossover, blocks that carry hard pins travel intact. Mutation inside a hard-pinned bar is limited to ornamental edits and register shifts that preserve the contour and the rhythmic spine of the theme. When a soft window is active, selection favours individuals that realise the theme within the window while maintaining tonal continuity at the entry point. If a theme is repeatedly missed in the designated window, the controller increases the probability of choosing parent blocks that already contain a compatible fragment and reduces the chance of disruptive edits in the bars immediately before the window.

Recognition of realised themes uses compact symbolic embeddings trained on the same library that supplies the prototypes. The detector searches for the best match within targeted windows and across the intended voice set. A match updates schedule adherence statistics and may adjust the pressure on subsequent windows. If a theme enters earlier than planned but the entrance is musically convincing and does not harm later commitments, the policy layer can accept the deviation and shift the next entrance accordingly. This flexibility keeps the process musical rather than mechanical.

The policy layer also supervises how textures evolve across sections. The run begins in a lean texture with close voice spacing and moderate rhythmic density to stabilise tonality and establish the main motif. As the piece advances, density and spacing widen, countermelodies gain independence, and the scheduler allows brief overlaps of themes to create dialogue. Before major cadences, the controller reduces mutation intensity and protects cadence bars from being split so that the repair pass can polish a convincing close. The final section returns to a clarified texture that recalls the opening theme in a new register or with mild augmentation, a strategy that helps listeners perceive long-range unity.

Throughout the run, diagnostics inform small adjustments. If violation counts increase in a specific voice, range limits for that voice are tightened and the probability of leaps is reduced in the next generations. If tonal drift appears in transitional regions, the controller brings back chord-tone bias on strong ticks and limits off-scale colour tones until stability returns. If recognisability of a theme is high but the accompaniment repeatedly obscures its rhythm, mutation is redirected toward rhythmic thinning in the other voices rather than further edits to the theme voice. These targeted reactions allow the system to correct course without suppressing the creative potential of the search.

This chapter has described the working logic of the adaptive genetic framework without recourse to formulas or diagrams. The design unites a bar-level representation that respects musical structure, a set of domain-aware operators, a fast local repair routine, and a policy layer that manages thematic commitments. The combination yields a search process that is disciplined enough to maintain contrapuntal hygiene and tonal direction, yet flexible enough to support variation, overlap, and long-range development across multiple themes.

4 Experimental results and analyses

4.1 Setup and evaluation protocol

- *Corpora and splits:* Experiments used three symbolic corpora curated for polyphonic writing: a chorale-style set with tight voice-leading, a chamber subset with duet and trio textures, and a modern tonal miniature collection. Each corpus was partitioned into non-overlapping training and evaluation pools. The training pool supplied the bar-aligned block library, theme prototypes, and the style model used only for scoring. The evaluation pool provided source material for theme scheduling and held-out harmonic profiles. No bar in the evaluation pool appeared in the block library. Corpora consist of curated chorale, chamber, and modern tonal sets under research friendly licenses with citation requirements. Splits use non overlapping pools for training and evaluation. Blocks for the library come only from training data. Theme prototypes and style scorers never access evaluation bars, preventing leakage across experiments.

- *Baselines:* Comparisons included a classic genetic algorithm with fixed operator rates, a particle-swarm heuristic configured for symbolic sequences, and a transformer language model followed by a post-hoc rule filter. All systems produced scores with the same number of voices, bar length, and target duration. Each configuration ran with five different random seeds.
- *Metrics:* Objective analysis covered contrapuntal violations per 100 beats, vertical consonance, accented tonal stability, theme recognisability, and population diversity. Runtime was reported as seconds per 1,000 evaluated candidates. For human evaluation, a listening panel rated perceived coherence and theme recognisability on a five-point Likert scale using double-blind presentation of short excerpts rendered from the scores. Contrapuntal violations per hundred beats capture voice crossing hidden parallels accented dissonances and unidiomatic leaps. Vertical consonance and accented tonal stability reflect harmonic soundness and key anchoring. Theme recognisability measures detector matches within scheduled windows. Diversity summarises structural variety. Scores average over excerpts then over seeds with bootstrap confidence intervals.
- *Procedure:* For every seed and method, the evolutionary loop ran until any of the following conditions held: fitness stabilised within a narrow band, mandatory theme entrances were realised with acceptable recognisability and spacing, or the generation budget was reached. The best individual per run was exported as MIDI and as a staff-notated score. Diagnostics recorded violation counts by category, tonal drift indices, and convergence traces, which were later aggregated for analysis.

4.2 Thematic control and policy layer

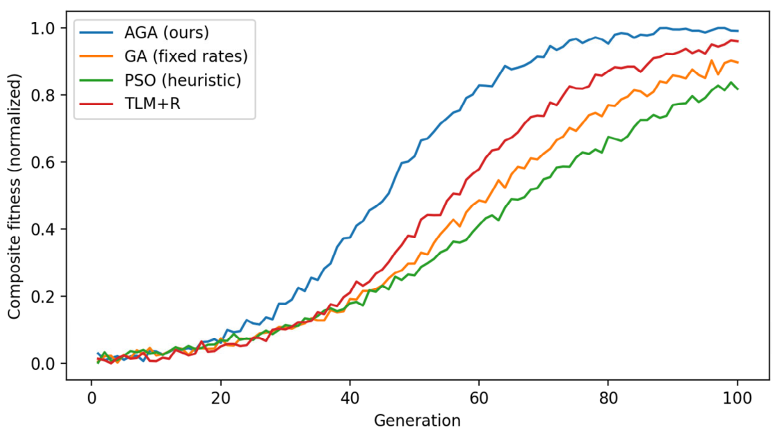
Table 1 summarises objective metrics averaged over the five seeds. Lower violation rates and higher consonance, tonal stability, and recognisability indicate better performance. The adaptive approach reduces infractions by a substantial margin while simultaneously improving harmony and theme clarity. Population diversity remains markedly higher, which aligns with the intention of delaying premature convergence. Runtime per 1,000 candidates remains in the same range as the fixed-rate GA and PSO. The transformer baseline incurs additional time in sampling and filtering. Reported gains include uncertainty estimates through ninety five percent bootstrap confidence intervals and effect sizes for key metrics. Hypotheses are evaluated using paired Wilcoxon signed rank tests with two sided thresholds at p less than 0.05 and false discovery control by Benjamini-Hochberg across families of comparisons.

The fitness trajectory confirms faster ascent and earlier saturation for the adaptive search. In early generations, diversity-aware crossover accelerates structural exploration without collapsing to clones. In the middle stage, mutation intensity rises briefly when stagnation is detected, which injects rhythmic and tonal alternatives that are then consolidated by selection and local repair. The fixed-rate GA tracks a slower curve and typically requires more generations to reach the same band. The transformer baseline climbs more quickly than PSO but flattens earlier, reflecting the ceiling imposed by its post-hoc filtering. To make the dynamics of training visible, Figure 2 charts the normalised composite fitness over 100 generations for all methods, averaged across five random seeds. The adaptive genetic approach climbs rapidly during the early and mid stages and settles at a higher plateau, indicating faster discovery of viable structures and more stable consolidation. The fixed-rate GA and PSO improve more gradually and require additional generations to reach the same band, while the transformer with rule filtering rises quickly at first but levels off earlier, suggesting a ceiling imposed by post-hoc corrections. Taken together with the objective scores in Table 1, these trajectories show that diversity-aware crossover, targeted mutation, and lightweight repair convert exploration effort into durable gains rather than short-lived spikes.

Table 1 Objective metrics across methods

Task no.	*Method*	*Contrapuntal violations*	*Vertical consonance*	*Accented tonal stability*	*Theme recognisability*	*Population diversity*
1	AGA	2.10 ± 0.50	0.83 ± 0.02	0.88 ± 0.02	0.79 ± 0.03	0.62 ± 0.04
2	GA	5.80 ± 1.10	0.77 ± 0.03	0.82 ± 0.03	0.61 ± 0.05	0.41 ± 0.06
3	PSO	6.40 ± 1.20	0.75 ± 0.03	0.80 ± 0.03	0.57 ± 0.05	0.38 ± 0.06
4	TLM + R	3.90 ± 0.80	0.80 ± 0.02	0.85 ± 0.02	0.73 ± 0.04	0.29 ± 0.05

Figure 2 Convergence profiles (see online version for colours)



4.3 Ablations and listening study

Table 2 isolates the contribution of adaptive control, local repair, and the topic scheduler. Removing adaptive control slows progress and increases infractions, indicating that rate updates guided by diversity and stagnation make exploration more productive. Disabling the repair pass has the most pronounced effect on violation counts and delays stabilisation, as illegal intervals accumulate and must be corrected by chance. Excluding the scheduler leaves much of the harmonic and contrapuntal quality intact but sharply reduces recognisability, confirming that explicit planning is essential for multi-topic writing.

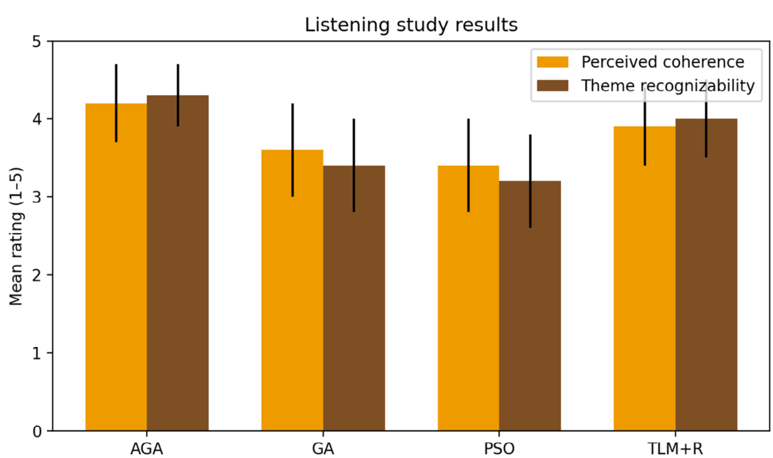
Twenty-four listeners with mixed formal training evaluated eight excerpts per method in randomised order. Each excerpt lasted ten to fifteen seconds and contained at least one thematic entrance. Participants rated perceived coherence and the ease of recognising themes. The adaptive system scored the highest on both scales, with the strongest margin on recognisability. The transformer baseline received competitive ratings on theme clarity but slightly lower coherence, consistent with the tendency of its filter to enforce local rules while missing longer-range dependencies. PSO and the fixed-rate GA trailed in both categories. The listening study enrolled twenty four participants with varied musical training backgrounds. Excerpts of ten to fifteen seconds were rendered from staff notation and randomised under double blind presentation. Order effects were countered by Latin squares. Agreement and consistency were summarised by Kendall's W with bootstrapped intervals.

Table 2 Ablation of key components

Task no.	*Configuration*	*Violations*	*Vertical consonance*	*Theme recognisability*	*Generations to target*
1	Full AGA	2.10	0.83	0.79	42
2	No adaptive control	3.20	0.81	0.73	61
3	No repair	4.70	0.79	0.74	68
4	No topic scheduler	3.60	0.82	0.41	55

Figure 3 shows mean listener ratings (1~5) for perceived coherence and theme recognisability across methods; error bars indicate one standard deviation. The adaptive genetic approach scores highest on both criteria, TLM+R ranks second, and GA/PSO trail.

Figure 3 Mean listener ratings for perceived coherence and theme recognisability across methods (see online version for colours)



Exemplars from the adaptive system display smoother inter-voice motion near cadences and more consistent placement of chord tones on metrically strong ticks. The scheduler supports clear entrances of multiple themes without suppressing development; overlaps occur in a way that listeners describe as dialog rather than clutter. When violations do appear, they are shorter and more localised, which aligns with the design of the memetic pass that targets exposed regions. In contrast, the fixed-rate GA exhibits episodes of premature uniformity: large portions of a piece converge toward similar rhythmic templates, after which mutations either fail to escape or introduce jarring edits that disrupt harmonic flow. The transformer baseline maintains idiomatic surface patterns but occasionally drifts in tonal direction across several bars, especially when a planned theme entrance requires a change in register or rhythm.

5 Conclusions

This study introduced an adaptive genetic framework for multi-topic polyphonic music generation that treats a score as bar-aligned, voice-aware chromosomes enriched with tonal and thematic metadata. The pipeline couples careful pre-processing and a topic scheduler with a domain-aware evolutionary core: voice-preserving crossover, musically constrained mutation, and a lightweight memetic repair focused on exposed positions such as cadential windows. An evaluator aggregates contrapuntal legality, harmonic soundness, accented tonal stability, thematic recognisability, and population diversity, while an adaptive controller adjusts operator rates in response to diversity and stagnation signals.

Across chorale, chamber, and modern tonal corpora, the approach consistently lowered rule violations and raised consonance and tonal alignment, and it improved theme clarity without inflating runtime. Convergence profiles showed earlier and higher plateaus than fixed-rate search or heuristic alternatives, and listening tests favoured the produced excerpts for coherence and recognisability. Ablations confirmed that each ingredient matters: adaptive control prevented premature uniformity, local repair suppressed error accumulation, and explicit scheduling proved decisive for multi-topic writing.

Several limitations remain. Expressive timing and dynamics were handled implicitly through metrical features rather than explicitly modelled performance layers; long-range form beyond bar-level blocks could be strengthened; and stylistic control relied on a frozen scorer rather than a learnable surrogate. Future work will extend the representation to phrase-scale structure and dynamics, introduce differentiable or learned fitness proxies for faster search, and support interactive use where a composer can lock, replace, or reshape sections mid-run. Despite these open avenues, the results indicate that adaptive evolutionary search, when grounded in polyphonic constraints and thematic planning, can produce structured, stylistically credible music while keeping the generation process transparent and controllable.

Acknowledgements

This work is supported by the Qingyuan City 25th Batch (2024) Basic Education Research Project named: Symbiotic Integration of Vocational Aesthetic Education and Yao Ethnic Intangible Cultural Heritage in Liannan under the Information Age (No. 25-97).

Declarations

All authors declare that they have no conflicts of interest.

References

Abbasi, S., Rahmani, A.M., Balador, A. and Sahafi, A. (2023) 'A fault-tolerant adaptive genetic algorithm for service scheduling in internet of vehicles', *Applied Soft Computing*, Vol. 143, p.110413.

Acharya, U.K. and Kumar, S. (2021) 'Genetic algorithm based adaptive histogram equalization (GAAHE) technique for medical image enhancement', *Optik*, Vol. 230, p.166273.

Ali, M.Z., Awad, N.H., Suganthan, P.N., Shatnawi, A.M. and Reynolds, R.G. (2018) 'An improved class of real-coded genetic algorithms for numerical optimization☆', *Neurocomputing*, Vol. 275, pp.155–166.

Hao, K., Zhao, J., Li, Z., Liu, Y. and Zhao, L. (2022) 'Dynamic path planning of a three-dimensional underwater AUV based on an adaptive genetic algorithm', *Ocean Engineering*, Vol. 263, p.112421.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A-X., Chen, W., Kougias, I. and Kazakis, N. (2018) 'Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution', *Science of the Total Environment*, Vol. 621, pp.1124–1141.

Ju, X. and Liu, F. (2019) 'Wind farm layout optimization using self-informed genetic algorithm with information guided exploitation', *Applied Energy*, Vol. 248, pp.429–445.

Kaliakatsos-Papakostas, M., Floros, A. and Vrahatis, M.N. (2020) 'Artificial intelligence methods for music generation: a review and future perspectives', *Nature-Inspired Computation and Swarm Intelligence*, pp.217–245.

Kuptametee, C., Michalopoulou, Z-H. and Aunsri, N. (2023) 'Adaptive genetic algorithm-based particle herding scheme for mitigating particle impoverishment', *Measurement*, Vol. 214, p.112785.

Luo, X. and Oyedele, L.O. (2021) 'Forecasting building energy consumption: adaptive long-short term memory neural networks driven by genetic algorithm', *Advanced Engineering Informatics*, Vol. 50, p.101357.

Luo, X., Oyedele, L.O., Ajayi, A.O., Akinade, O.O., Owolabi, H.A. and Ahmed, A. (2020) 'Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings', *Renewable and Sustainable Energy Reviews*, Vol. 131, p.109980.

Ma, C. and He, R. (2019) 'Green wave traffic control system optimization based on adaptive genetic-artificial fish swarm algorithm', *Neural Computing and Applications*, Vol. 31, No. 7, pp.2073–2083.

Ma, J., Gao, W. and Tong, W. (2025) 'A deep reinforcement learning assisted adaptive genetic algorithm for flexible job shop scheduling', *Engineering Applications of Artificial Intelligence*, Vol. 149, p.110447.

Mokshin, A.V., Mokshin, V.V. and Sharnin, L.M. (2019) 'Adaptive genetic algorithms used to analyze behavior of complex system', *Communications in Nonlinear Science and Numerical Simulation*, Vol. 71, pp.174–186.

Ness, I.J. and Dysthe, O. (2020) 'Polyphonic imagination: understanding idea generation in multidisciplinary groups as a multivoiced stimulation of fantasy', *Creative Learning in Digital and Virtual Environments*, pp.50–67, Routledge.

Prado, F., Minutolo, M.C. and Kristjanpoller, W. (2020) 'Forecasting based on an ensemble autoregressive moving average-adaptive neuro-fuzzy inference system–neural network-genetic algorithm framework', *Energy*, Vol. 197, p.117159.

Que, Y., Zhong, W., Chen, H., Chen, X. and Ji, X. (2018) 'Improved adaptive immune genetic algorithm for optimal QoS-aware service composition selection in cloud manufacturing', *The International Journal of Advanced Manufacturing Technology*, Vol. 96, No. 9, pp.4455–4465.

Resende, P.A.A. and Drummond, A.C. (2018) 'Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling', *Security and Privacy*, Vol. 1, No. 4, p.e36.

Xue, Y., Zhu, H., Liang, J. and Słowik, A. (2021) 'Adaptive crossover operator based multi-objective binary genetic algorithm for feature selection in classification', *Knowledge-Based Systems*, Vol. 227, p.107218.

Yadav, P.S., Khan, S., Singh, Y.V., Garg, P. and Singh, R.S. (2022) 'A lightweight deep learning-based approach for jazz music generation in MIDI format', *Computational Intelligence and Neuroscience*, No. 1, p.2140895.

Yuan, N. (2024) 'Does AI-assisted creation of polyphonic music increase academic motivation? The deepbach graphical model and its use in music education', *Journal of Computer Assisted Learning*, Vol. 40, No. 4, pp.1365–1372.

Zhang, H-L. and Yu, P-J. (2021) '[Retracted] design of multimedia vocal music education data integration system based on adaptive genetic algorithm', *Security and Communication Networks*, No. 1, p.3897696.

Zhang, J., Ding, L. and Long, S. (2024) 'State feedback control for vehicle electro-hydraulic braking systems based on adaptive genetic algorithm optimization', *International Journal of Intelligent Systems*, Vol. 2024, No. 1, p.3616505.

Zhou, M., Long, Y., Zhang, W., Pu, Q., Wang, Y., Nie, W. and He, W. (2021) 'Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization', *IEEE Transactions on Evolutionary Computation*, Vol. 25, No. 5, pp.913–927.