ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Construction of Vocal Rehabilitation Model and Personalized Treatment Pathways Driven by Multi-Objective Optimization Algorithm

Naqi Yan[1,*]

[1] College of Humanities and Arts, Guangxi Vocational University Of Agricultural, Nanning, Guangxi, 530000, China

SUMMARY: *The vocal professionals are prone to voice disorders as a result of excessive exposure to high intensity voices that can impact their job performance and the quality of their life. Standardized treatments often suffer from lack of multi-dimensional optimization and lack of personalization. The aim of this study is to overcome such shortcomings by offering a multi-objective optimization algorithm-based strategy of building a vocal rehabilitation model and personalized treatment pathways with the view to achieving synergistic optimization on various therapeutic dimensions. Based on the combination of acoustic properties, physiologic measurements, and subjective perception ratings, a multi-objective optimization model is constructed. An improved Non-dominated Sorting Genetic Algorithm II (NSGA-II) is created to effectively search through the set of Pareto-optimal solutions to the treatment parameters. The experimental evidence shows that the proposed solution is much more effective than the traditional single-objective optimization and manual empirical solutions in terms of reducing the duration of the treatment and increasing patient comfort as well as ensuring high accuracy in acoustic restoration. These findings offer a new data-driven framework in personalized medical care in voice rehabilitation and extend the use of multi-objective optimization algorithms in biomedical engineering.*

KEYWORDS: *voice rehabilitation; personalized treatment pathway; multi-objective optimization; improved non-dominated sorting genetic algorithm; intelligent rehabilitation*

1 Introduction

Vocal professionals sometimes suffer from voice disorders because of excessive and intense vocal usage, which heavily disrupts their work and quality of life [1]. Despite the existence of different treatment options, traditional ones tend to be symptomatic or empirically oriented and do not involve a comprehensive consideration of multi-dimensional therapeutic effects. It is this weakness that renders them unfit to handle the significant differences between individuals regarding their physiological build, acoustic properties, and personal experiences of sound, leading to prolonged rehabilitation times, heavy treatment loads, and poor patient satisfaction [2]. There is an urgent need to create an adaptive optimization system, which will include the acoustic, physiological, and perceptual aspects into consideration, thus making the treatment strategies more individualized and synergistic. The multi-objective optimization method proposed in this paper involves the application of a multimodal evaluation framework to come up with an intelligent voice restoration model as a solution to voice disorders.

---

*yannaqi85@163.com
https://doi.org/10.65102/is2026465

Moreover, it examines the best combinations of therapeutic parameters to attain better multi-dimensional rehabilitation results.

In recent years, voice disorder studies have attracted more consideration in medical engineering and rehabilitation sciences. Previous works have used speech signal processing technology to obtain acoustic properties and statistical modeling algorithms to detect abnormal vocal patterns as part of voice condition surveillance and prognosis prediction [3, 4]. In terms of intervention strategies, the traditional training protocol and the behavioral modification technique are still being improved. Other researchers have investigated feedback-regulated vocal training devices, which enhance vocal performance and adherence to treatment in some patients [5, 6]. At the same time, the use of artificial intelligence technologies to recognize speech and lesions has brought about new opportunities in the fields of intelligent rehabilitation pathways [7, 8]. However, even with the current level of technological growth and the creation of evaluation systems in it, most of the existing approaches are limited to diagnostic assistance or local measures and do not incorporate systematically therapeutic factors and multidisciplinary response systems. Professional voice users or patients with high vocal demands are frequently treated through an empirical approach to planning their treatments, and there are no established data-driven systems and dynamism to meet the requirements of exact and individualized care. To overcome these constraints, a multimodal data-driven optimization paradigm will be suggested that consists of simulating and optimizing vocal rehabilitation programs by utilizing a more advanced multi-objective evolutionary algorithm to balance diverse treatment objectives dynamically. Experiments have shown that this approach can enhance general treatment efficacy considerably, and it is superior to traditional approaches in quite a number of ways. They both validate the value and effectiveness of multi-objective optimization as applied to medicine rehabilitation, and provide a systematic and data-driven model of the voice disorder intervention strategies. Its theoretical contribution increases the structural representation of the therapeutic mechanisms and its practical contribution extends the design boundaries of individualized adaptive treatment protocols.

The chapter structure is as follows: The first chapter gives a background to the research, problems, and importance; The second chapter has a review of the existing advances in the area of multi-objective optimization algorithms in medical rehabilitation; The third chapter describes the data sources, evaluation models, and construction of an algorithm; The fourth chapter displays and evaluates the model performance and optimal rehabilitation pathways; The fifth chapter reflects and analyzes the results; The sixth chapter ends the research and offers the direction of future research.

2 Literature Review

The multi-objective optimization (MOO) algorithms have gained significant attention in the field of rehabilitation engineering due to their unique advantages in addressing conflicting objectives and improving the performance of the system [9]. Unlike conventional single-objective methods, MOO allows establishing Pareto-optimal solutions in several dimensions, such as training efficiency, energy usage, safety, and individual flexibility. The ability offers a more flexible and generalizable model of dealing with complicated issues like rehabilitation pathway planning, control strategy design, and intervention on functional impairments [10, 11]. Since the combination of computational intelligence and sensing technologies, the use of MOO increased in robot-assisted training systems, as well as in therapies of voice and swallowing disorders. Such a trend can be clearly seen when it comes to the process of solution integration, which moves toward personalized optimization due to the development of adaptive therapeutic design.

When designing a rehabilitation path, goals are commonly composed of several factors, such as training effectiveness, smoothness of the path, energy consumption, and personal differences. The introduction of MOO has changed the process of path generation as it is no longer aimed at finding a single optimal solution but rather to create a set of Pareto-optimal solutions that will meet the needs of different patients. NSGA-III was used by Yu et al., (2021), to optimize tri-objectives high-degree-of-freedom redundant rehabilitation robot trajectories, and showed the algorithm to have a synergistic ability in the balance of time efficiency, stability, and operational accuracy in complex systems [12]. Particle swarm optimization was then combined with bat algorithm by Abed and Jasim (2022) to effectively balance both obstacle avoidance and energy-efficient path optimization [13]. Fan et al. (2023) created a multi-objective trajectory planning model to be used in upper-limb rehabilitation contexts where accessibility, training intensity and muscle loading were optimized simultaneously to offer customized support to patients on their rehabilitation objectives [14]. Also, Nickelson et al. (2023) suggested the Contextual Multi-Objective Path Planning (CMOPP) algorithm that allows separating the modeling of the cost of a path and its generation process to effectively decrease the size of the search space [15]. This method allows dynamically coordinating the conflicting goals like path length, safety, and obstacle avoidance having great flexibility in the design of personalized rehabilitation robots paths. In order to apply it to more general conditions, Wang et al. (2025) extended multi-objective path planning to dynamic urban traffic environments [16]. Through its inclusion of dynamic weighting schemes, they achieved a balanced optimization of different objectives which include path length, energy consumption, safety and congestion levels hence forming a very extensible theoretical framework of multi-scenario rehabilitation navigation system. The MOO algorithms have been effectively applied to eliminate the disadvantages of the idea of fixed criteria involved in conventional path design approaches. Such methods let rehabilitation training pathways react dynamically to the individual differences and multi-objective factors and user feedback, and are hence very scalable and generalizable.

Robot control systems that are applied in rehabilitation should consider execution accuracy, energy consumption, safety, response time and their interaction with humans. MOO provides a coordinating device on such interrelated yet at times conflicting goals, particularly applicable to complex dynamic problems. An adaptive weight allocation procedure based on deep reinforcement learning has been proposed by Wang and Xu (2022) according to which the order of objectives can be dynamically adjusted, which allowed them to develop control algorithms that could address different steps of recovery [17]. According to Widanage et al. (2023), the demand-oriented control strategy based on MOO was proposed to balance the level of patient involvement and intensity of assistance to achieve a high level of stability in gait and user comfort in Parkinsonian patients [18]. Using the Kriging surrogate model, the authors Li et al. (2023) optimized stiffness and weight when creating lower-limb support structures by minimizing the power consumption of the devices without affecting the stability of the structure [19]. Physiological feedback in real-time was done through electromyography (EMG) signals when implementing MOO-based control strategies to enhance responsiveness and prove superior performance under complicated control conditions [20]. MOO algorithms enable transferring robotic control strategies between standardised templates and personalized adaptation increasing precision and interactivity of rehabilitation training. This is one of the major algorithmic tendencies of intelligent rehabilitation robot systems.

The goals of the voice and swallowing disorders rehabilitation go further than clinical measures of swallowing efficiency, speech clarity, and respiratory coordination into the subjective aspects of cognitive load, compliance to training, and quality of life. MOO allows systematic alignment of these diverse objectives. The review article by Matos et al. (2022) indicated that although various methods are used together, there are no systematic optimization standards [21]. MOO offers a unified system of optimizing the choice of intervention strategies, training intensity, and frequency. The speech-language pathology (SLP) survey by Chen et al. (2022) indicated that clinical practice has been limited by limited resources and goal conflicts, which is why it is necessary to use MOO evaluation when developing the standard [22]. In 2023, Zhang introduced the concept of speech-visual co-intervention, which illustrates the principles of MOO through enhancing the language expression but minimizing the cognitive load [23]. Liu et al. (2024) found that patients who undergo surgery to treat head and neck cancer commonly suffer from both speech and swallowing impairment at the same time, and require a balance between structural restoration and functional recovery in the early treatment [24]. In addition, Srinivasan et al. (2023) pointed out that machine learning with MOO solutions can be very useful to learn the coupling interactions between swallowing function and speech properties, especially in the prediction after surgery and remote diagnosis [25].

The available literature has thoroughly examined how multi-objective optimization (MOO) algorithms have been used in rehabilitation engineering, specifically creation of an overall analysis approach of path planning, control policy scheduling and multi-dimensional therapeutic efficacy balance, and thus it creates theoretical and methodological basis of the present research. Nevertheless, most studies today concentrate on limb rehabilitation or generalized intervention tasks, and frequently ignore voice disorders, a rehabilitation issue that is highly dependent on multimodal assessment systems, and rarely consider the interrelated processes between subjective perception, physiological load, and acoustic performance. To bridge these gaps, the suggested research recommends a customized optimization model that has cross-modal indicators. To address the issue, an improved NSGA-II algorithm is created that will dynamically generate Pareto-optimal solution sets that intend to develop an intelligence-driven rehabilitation pathway system capable of adjusting to the diverse characteristics of vocal impairment and achieving synergy results in more than one therapeutic domain.

3 Research Design

3.1 Dataset and Experimental Configuration

The data set that is utilized in this paper has been jointly developed by our research team alongside the Otolaryngology Department of a tertiary hospital that contains the acoustic, clinical, and subjective perceptual data of 40 professional vocal performers. The subjects consisted of 25 people with functional voice disorders and 15 normal controls, aged 25-52, and had a minimum of three years of professional vocal training. Ethical measures were strictly enforced and the participants did not have any acute disease, such as colds or pharyngitis, on the day of recording and written consent was obtained. Voice recording was performed in an anechoic chamber and recorded using a condenser microphone (44.1 kHz sampling rate, 16-bit resolution) and recorded using a professional acoustic analysis system. Standard vocal tasks were 5 seconds sustained vowel phonation, standardized sentence reading and spontaneous speech. All participants gave three vocal samples which resulted in 120 high fidelity recordings. In order to ensure that acoustic, physiological and subjective information were matched across time in all three areas, physiological measurements, such as laryngoscopic video, laryngeal electromyography (EMG) and spirometric data, were measured at the same time. Training, validation and test sets were divided in the ratio of 7:2:1, and the five fold cross-validation was employed in order to increase the generalizability of the model.

The experiments were conducted on a supercomputer platform under the Windows 11 operating system, which has an Intel Core i9-13900K CPU, 32 GB memory and an NVIDIA RTX 3080 graphics card. The entire preprocessing operations of the data were completed in the Python 3.10 environment, mostly in these steps: Background noise was removed and the energy was normalized by applying a wavelet threshold-based denoising algorithm to the speech signals. Then, the speech signals were split into frames using a sliding window approach with a frame length of 25ms and a frame shift of 10ms, and the Hamming window function was applied to every frame. The obtained frame-level acoustic features were normalized to fixed-dimensional feature vectors. Physiological data were laryngeal EMG signals filtered with a band-pass filter (20-500Hz) and normalized to z-scores. Laryngoscopic images were processed by image enhancement and region of interest (ROI) extraction method in which major anatomical regions were coded as constant structural index. Pulmonary function parameters were incorporated as model inputs following missing value inputation and unit normalization. Objective and subjective rating data were converted by converting negative item scores to positive and scaling them to a 0-100 scale, which was then transformed and combined with objective variables. It produced a common feature base to serve in multimodal modeling and optimization algorithm applications of the future.

3.2 Construction of Cross-Modal Evaluation Framework

This paper has developed a cross-modal assessment structure that is based on three dimensions of speech rehabilitation results to thoroughly evaluate the voice rehabilitation results and be able to offer credible support in multi-objective optimization. It combines such important elements as the status of the phonatory stability, the condition of the glottal function, and self-perception of the patients, which makes sure that the optimized treatment pathways lead to synergistic improvements of objective performance and personal experience. There are several metrics included in each dimension and how they were obtained explained in Table 1.

Table 1: Cross-modal Voice Rehabilitation Evaluation Index System

Dimension	Evaluation Index	Definition	Measurement Method/Tool
Acoustic Features	Jitter	Reflects the instability of vocal fold vibration period (%)	Praat software for speech signal analysis
	Shimmer	Quantifies glottal closure intensity and amplitude fluctuation (%)	MDVP software automatic calculation
	Fundamental Frequency Standard Deviation (F0 SD)	Describes the variation range of vocal pitch (Hz)	Speech spectrum extraction tools
	Glottal Closure Index Ratio (GCI Ratio)	Measures the completeness of glottal cycle closure (0~1)	Glottal cycle analysis tools
Physiological Parameters	Maximum Phonation Time (MPT)	Maximum sustained phonation duration, reflecting pulmonary function and glottal coordination (s)	Timer + recording device
	Subglottal Pressure (SP)	Airflow pressure below glottis, indicating vocal driving force (cmH_2O)	Air pressure measurement device
	Vocal Fold Impedance (VFI)	Glottal impedance variation during phonation (normalized)	High-speed imaging & computational modeling system
Subjective Perception	Hoarseness Score (HS)	Characterizes subjective abnormal vocal qualities (e.g., roughness, tension, weakness)	GRBAS scale
	Voice Handicap (VH)	Patient's subjective perception of voice disorder's impact on daily life	VHI-10 scale

3.3 Construction of Multi-Objective Optimization Model

In order to realize the collaborative optimization of personalized treatment goals of vocal disorders rehabilitation, this paper created a multi-objective optimization model that was rooted in the cross-modal evaluation framework. The model includes acoustic recovery accuracy, physiological coordination and subjective comfort with an objective to improve speech feature stability and efficiency of physiological phonatory mechanisms as well as the level of patient satisfaction across various combinations of treatment parameters, thus producing personalized optimal pathways dynamically.

Specifically, the treatment parameter combination is defined as the decision variable vector $\mathbf{x}=[x_1,x_2,\ldots,x_n]$, which includes intervention components such as vocal fold movement training intensity, respiratory support training duration, and glottal control feedback frequency. The model consists of three objective functions forming a non-dominated optimization system, as follows:

$f_1(\mathbf{x})$: Minimize the comprehensive score of acoustic perturbation indices (e.g., Jitter, Shimmer, F0 SD);

$f_2(\mathbf{x})$: Minimize physiological energy expenditure indices (e.g., Subglottal Pressure, Vocal Fold Impedance);

$f_3(\mathbf{x})$: Maximize subjective satisfaction scores (e.g., inverse scores of Voice Handicap Index, Hoarseness subjective ratings).

Here, $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are objectives to be minimized, while $f_3(\mathbf{x})$ is to be maximized. To align the optimization direction, this study applies a negative transformation to $f_3(\mathbf{x})$, converting all objective functions into minimization problems:

$$\min_{\mathbf{x}\in\mathcal{X}}\left[f_1(\mathbf{x}), f_2(\mathbf{x}), -f_3(\mathbf{x})\right] \tag{1}$$

where $\mathcal{X}$ represents the feasible solution space for all treatment parameter combinations, constrained by practical training intervention conditions and individual physiological response thresholds.

3.4 Design of the Improved NSGA-II Optimization Algorithm

The current paper presents a multi-objective optimization algorithm that is grounded on the Non-dominated Sorting Genetic Algorithm II (NSGA-II) algorithm that is specific to voice rehabilitation problems (NSGA-VR). This model-solving central driver incorporates three dimensions of evaluation, namely, acoustic perturbation, physiological load, and subjective perception in one evolutionary optimization scheme with elitism, adaptive mutation, and diversity control (Figure 1). Construction step-by-step is:

(1) Individual Encoding and Initial Population Generation

Every solution vector (that is, a treatment parameter combination within the model) is indicated as $\mathbf{x}=[x_1,x_2,...,x_n]$, which includes parameters like vocal training duration, phonation intensity adjustment coefficient, and glottal closure control frequency. Population initialization Population initialization is improved by using Latin Hypercube Sampling (LHS) to generate $N=100$ initial individuals that can cover more of the solution space with greater representativeness and diversity of parameter distribution on all dimensions.

(2) Fitness Evaluation and Non-dominated Sorting Mechanism

Each solution vector $\mathbf{x}$ has an objective function value computed using the cross-modal feature vectors extracted in Section 3.2, which yields normalized values of acoustic perturbation $f_1(\mathbf{x})$, physiological load $f_2(\mathbf{x})$ and subjective satisfaction $f_3(\mathbf{x})$, all being minimization problems. The non-domination relationship among individuals is given by:

$$\mathbf{x}_i \prec \mathbf{x}_j \Leftrightarrow \forall k,\ f_k(\mathbf{x}_i) \le f_k(\mathbf{x}_j) \wedge \exists k,\ f_k(\mathbf{x}_i) < f_k(\mathbf{x}_j) \tag{2}$$

Based on this relationship, fast non-dominated sorting is performed to partition the solutions into different levels $\mathcal{F}_1, \mathcal{F}_2, \ldots$. Within the same level, further diversity evaluation and ranking are conducted using the crowding distance (Crowding Distance):

$$CD_i = \sum_{k=1}^{m} \frac{f_k^{(i+1)} - f_k^{(i-1)}}{f_k^{\max} - f_k^{\min}} \tag{3}$$

where $f_k^{(i)}$ denotes the objective function value of the i-th ranked individual on the k-th objective.

(3) Adaptive Crossover and Mutation Strategy

To achieve dynamic control of the search process, this study introduces an adaptive probability adjustment mechanism based on the standard Simulated Binary Crossover (SBX) and polynomial mutation. This mechanism modulates the crossover probability p_c and mutation probability p_m, enabling gradual adjustments with iteration progression:

$$p_c(g) = p_{c0} \cdot \left(1 - \frac{g}{G_{\max}}\right)^{\alpha}, \quad p_m(g) = p_{m0} \cdot \left(1 - \frac{g}{G_{\max}}\right)^{\beta} \tag{4}$$

where g represents the current iteration count, and $G_{\max}$ denotes the maximum number of iterations.

(4) Termination Criterion and Optimal Solution Set Output

The algorithm employs the maximum iteration count $G_{\max}$ as the primary termination criterion. If the crowding distance variation $\Delta CD < 0.001$ of the non-dominated solution set in all three objective dimensions does not change over 10 consecutive generations, the evolution is stopped prematurely. The last output is the Pareto optimal solution set $\mathcal{P}^*$ that comprises various therapeutic parameter combination schemes. Each scheme is an intervention pathway with various trade-off strategies concerning acoustic accuracy, physiological coordination, and subjective experience and hence serves as the foundation on which customized treatment plans can be developed.

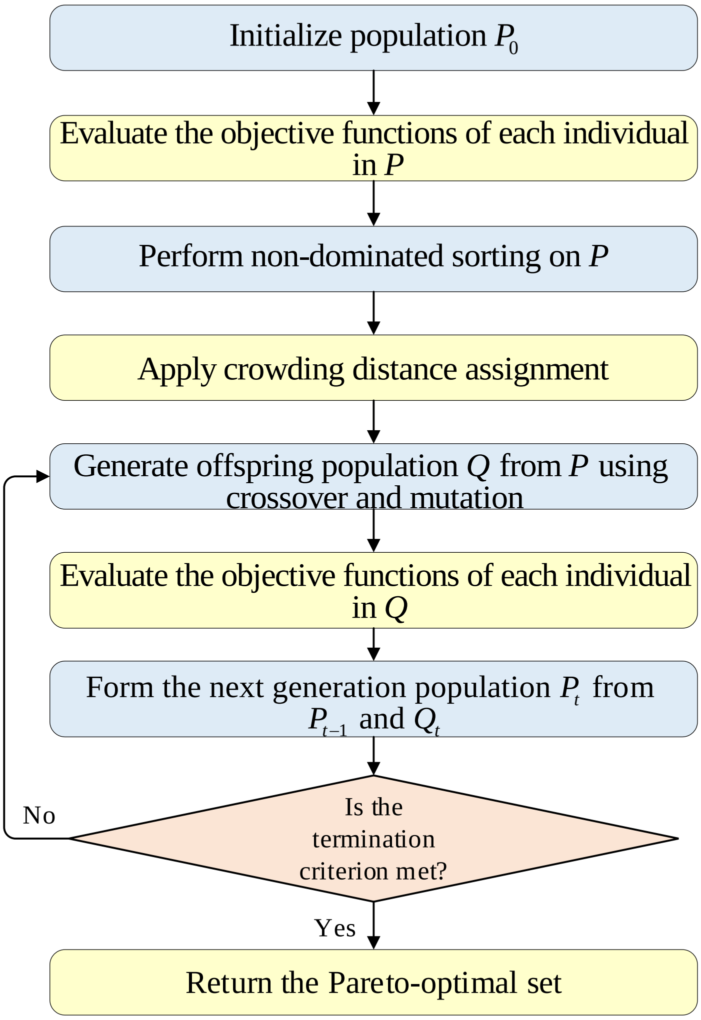

Figure 1: Framework diagram of the NSGA-VR model

3.5 Evaluation Metrics

To comprehensively evaluate the multidimensional performance of the proposed vocal rehabilitation model, five evaluation metrics were carefully selected across three dimensions: optimization quality, prediction accuracy, and user experience. These metrics cover critical aspects such as model convergence, numerical fitting capability, and patient physiological-behavioral responses.

Hypervolume (HV) is employed to assess the overall superiority and coverage capability of the Pareto front solution set. A larger HV value indicates that the solution set occupies a more optimal region in the objective space. It is mathematically defined as follows:

$$HV = \mathrm{Vol}\left(\bigcup_{x \in P} \prod_{i=1}^{m} [\, f_i(x), r_i \,] \right) \tag{5}$$

where P denotes the non-dominated solution set and r_i represents the reference point for each objective.

Spread (Δ) measures how evenly distributed are non-dominated solutions on the Pareto front, such that a lower Δ implies more balanced solution sets. Spread is defined as:

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{n-1} |d_i - \bar{d}\,|}{d_f + d_l + (n-1)\bar{d}} \tag{6}$$

where d_f and d_l denote the boundary distances, and $\bar{d}$ is the average inter-solution spacing.

The Mean Squared Error (MSE) assesses the predictive accuracy of a model on acoustic and physiological measures, where lower MSE values mean better fitting results. This can be expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

where y_i denotes the true observed value of the i-th sample, and $\hat{y}_i$ represents the model's predicted value for the i-th sample.

Variation of the intensity of facial electromyography response (ΔEMG) represents the level of tension alleviation in the muscles involved in phonation. It is measured with the help of the mean potential changes before and after therapy of main facial electromyography (EMG) recording points (e.g., thyrohyoid muscle):

$$\Delta EMG = \frac{EMG_{\text{pre}} - EMG_{\text{post}}}{EMG_{\text{pre}}} \times 100\% \tag{8}$$

The reduction of ΔEMG values is evidence of decreased muscle tension and more natural phonation after the treatment.

Change Rate of Task Completion Time (ΔTCT) is a measure of how the patient responds to training tasks by assessing the change in mean time taken to complete standardized phonation tasks before and after treatment:

$$\Delta TCT = \frac{TCT_{\text{pre}} - TCT_{\text{post}}}{TCT_{\text{pre}}} \times 100\% \tag{9}$$

The improved efficiency of task execution and efficacy of the therapeutic path are indicated by a higher ΔTCT.

4 Research Results and Analysis

4.1 Model Performance Analysis

4.1.1 Convergence Analysis

This paper carried out a comparative examination of their convergence behavior based on the analysis of the performance of the improved multiobjective optimization algorithm NSGA-VR in finding parameters of the therapeutic pathway as compared to the classical NSGA-II algorithm, as demonstrated in Figure 2.The convergence curves show that NSGA-VR has a steeper descent tendency in the initial optimization stage implying that it is more effective in terms of global search and hence can quickly reach the Pareto optimal front.

With successive iterations, the convergence rate of NSGA-VR reduces steadily, and the loss value oscillates less until a steady-state is reached at a low level, indicating increased local search capability and convergence stability. On the other hand, whereas NSGA-II also displays a fast initial decrease, it shows variation in the middle and late phases and ends up with a slightly higher loss value than NSGA-VR. It means that NSGA-II has certain limitations in terms of solution stability and exactness of search granularity in comparison with NSGA-VR.

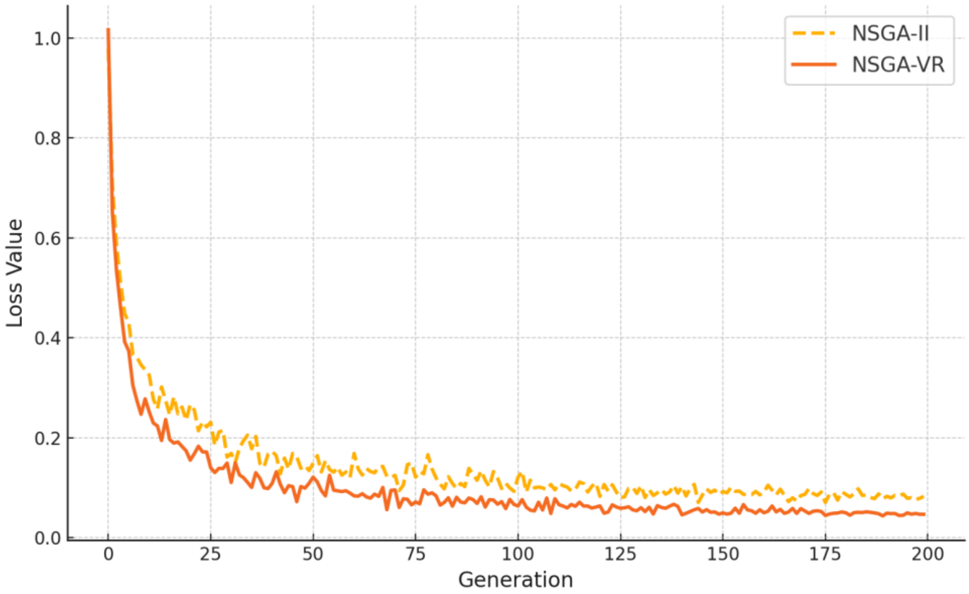

Figure 2: Convergence comparison between NSGA-VR and NSGA-II

4.1.2 Comprehensive Performance Comparison

This paper will compare NSGA-II, MOEA/D, and SPEA2 as comparative algorithms to systematically assess the general performance of NSGA-VR in multi-objective voice rehabilitation optimization problems. NSGA-II is a classical reference model which has been applied in most multi-objective optimization problems due to its use of crowding distance and a fast non-dominated sort algorithm. MOEA/D (Multi-Objective Evolutionary Algorithm based on Decomposition) reduces the multi-objective problem into a number of subproblems by using weight vectors and emphasizes structural guidance and local search capability in the solution set. SPEA2 (Strength Pareto Evolutionary Algorithm 2) is the implementation of fitness assignment with an external elitism mechanism whereby the strength of domination and the maintenance of diversity in the solution set are given consideration. These three algorithms are currently main frameworks of multi-objective optimization with large generality and great comparative value.

Figure 3 shows that NSGA-VR has the highest HV value (0.794), compared to NSGA-II (0.748), MOEA/D (0.768), and SPEA2 (0.752). It means that the non-dominated solution set produced by NSGA-VR covers more area in the objective space, indicating better overall solution quality and solution set variety. It confirms that NSGA-VR provides an efficient trade-off between exploring and converging, which offers a broader range of possible solutions to candidates to facilitate the development of personalized treatment plans. In terms of the Spread (Δ)) indicator, NSGA-VR was also the best (= 0.212), with lower average values than the rest of the algorithms, which implies that its solutions are more uniformly spread on the Pareto front and do not have too much clustering or empty spots in the solution set. As illustrated by these findings, NSGA-VR is more effective in ensuring the diversity of the solution sets, which enables them to offer a variety of treatment options to patients with different levels of vocal impairment.

In terms of model fitting accuracy, the MSE outcomes indicate that the NSGA-VR (0.008) had the lowest error of all models, with different levels of improvement over NSGA-II (0.011) and MOEA/D (0.010) and SPEA2 (0.009). This suggests that NSGA-VR has greater capability to reflect the nonlinear connections between acoustic and physiological features in the context of multi-objective optimization, which in turn increases the responsiveness of treatment pathways to real physiological feedback, and their forecasting power. When considering the Δ TCT indicator as an indicator of patient adaptation enhancement, NSGA-VR demonstrated the greatest increase (0.273), which means that the produced personalized pathways can greatly enhance efficiency of patients in executing standard vocal exercises. This can be explained by the fact that NSGA-VR attained a more balanced state between acoustic rehabilitation and physiological coordination. On the other hand, NSGA-II (0.194), MOEA/D (0.231), and SPEA2 (0.218) had significantly lower Δ TCT measures, which means that they have limited capabilities to increase patient compliance during the actual training.

Considering the variation in facial electromyographic response intensity (ΔEMG) as one of the most important physiological indicators indicating the state of muscle relaxation after treatment, NSGA-VR recorded the highest score (18.362 μV) which was significantly higher than that of NSGA-II (15.807 μV), MOEA/D (16.934 μV), and SPEA2 (17.083 μV). These findings indicate that NSGA-VR is more capable of controlling tension between the muscle groups involved in phonation, and it may be better able to reduce muscle tension associated with excessive vocal use, thus giving the patient a more natural and comfortable vocalization.

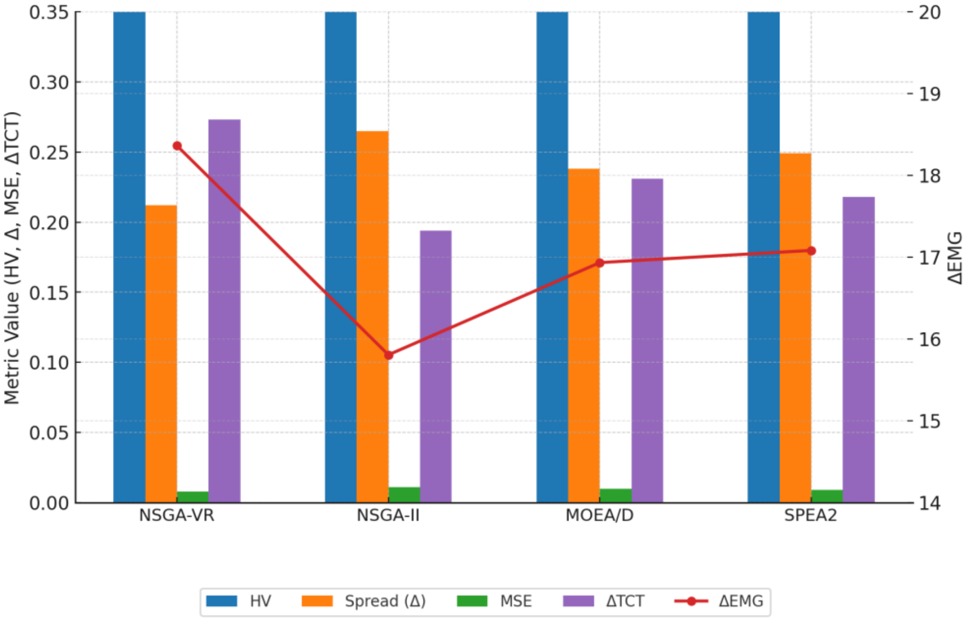

Figure 3: Comparative results of performance metrics

4.2 Solution Set Characteristics Analysis of Optimization Algorithms

To further validate the solution set performance of the improved NSGA-II algorithm (NSGA-VR) in voice rehabilitation parameter optimization, this study analyzes the distribution characteristics of the Pareto optimal solution set generated in the final iteration

(Fig. 4). Based on the general distribution, the Pareto solution set produced has good coverage and variety within the objective space with various non-dominated solutions spread out in other trade-off regions. It shows that the NSGA-VR algorithm has a high level of cooperative optimization in multi-objective optimization. In particular, a certain group of solutions has much smaller values in terms of f1 and f2 dimensions, which makes them appropriate to use in rehabilitation strategies focused on acoustic accuracy and physiological effectiveness. Another subset has lower subjective discomfort scores (f3 dimension) and therefore is more applicable to treatment goals focused on patient comfort perception.

Through further observation, it is found that there are different objective conflict properties in some boundary solutions: as they reduce acoustic perturbation (e.g., Jitter, Shimmer) they may cause a particular increase in subjective discomfort ratings. It indicates that some of the high-intensity vocal training pathways that are effective in enhancing acoustic performance need to be balanced with patient acceptability. Subsequently, there are particular solutions that provide better f3 objective scores at a marginal loss of acoustic accuracy and can be recommended to use as interventions when recovery is in its initial phases or when there is an excessive amount of vocal fatigue. Moreover, the solution set is also relatively evenly spaced across crowding distance without any severe clustering or collapse which leads to the conclusion that the NSGA-VR algorithm has significant power in conservation of solution diversity. There is a wide variety of possibilities on which further individualized treatment pathway suggestions can be made. On the whole, these characteristics of the solution set are highly confirming the usefulness of the multi-objective optimization framework to find trade-offs between voice therapy parameters space, providing solid data support to dynamically adaptive individualized intervention plans.

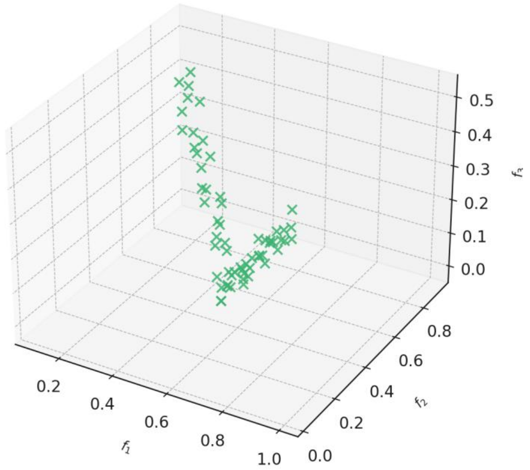

Figure 4: Three-dimensional distribution of Pareto optimal solutions

4.3 Dynamic Evolution Case Study of Personalized Treatment Pathway

In order to confirm the ability of dynamic adjustment of the optimization model in real-life rehabilitation, this paper has chosen a typical patient with moderate voice disorder as an exemplar and followed the development of the personalized pathway through 10 training sessions (Fig. 5). In the first stage (Sessions 1-3), the system suggested maintaining relatively mild pace of training, the duration of which was controlled at 15.2 to 16.7 minutes, and the frequency of glottal control was set at 95.4 to 99.6 Hz. At this phase, the model focused on the acceptance of training and psychological safety that were dependent on the high laryngeal muscle tension and poor psychological compliance of the patient during the first few sessions.

The subjective discomfort score went down by 0.66 to 0.61. Although the decrease was rather small, it indicated some early adaptation effects.

Mid-term stage (Sessions 47) has gone through a high parameter adjustment mode. The model, as it found the patient to have better glottal closure capability (as measured by the higher glottal closure index, GCI), and stable pulmonary function curves, progressively added more training load by increasing the duration of training by 16.3 minutes to 20.1 minutes and increased the glottal control frequency to 110.8 Hz. Following the 6th session, the subjective discomfort score decreased to 0.49 which means that there was a high effectiveness of intervention. Nevertheless, on the 7th session, the score returned to 0.52, and the model concluded that this was because of the training intensity being slightly higher than the optimal value of the individual, resulting in short-term tension. In turn, the system adjusted the glottal frequency of the 8th session automatically (114.6 Hz reduced to 112.9 Hz) in order to return the balance of experience.

In the later phase (Sessions 8-10) the model took a steady state adjustment approach, with training times being kept at over 20 minutes and frequencies at a high intensity level (>114 Hz). This was aimed at keeping the training effect on the phonatory system. At this point, the subjective score of the patient had reduced steadily to 0.35, indicating that the transition between the reliance on external instructions and active voice control had been successfully completed. The training path gradually evolved into an exploration-focused approach.

As was shown above, this evolution of the pathway represents the ability of the model to change dynamically the training regime on the fly using multi-dimensional feedback. Adjustment of parameters at all the stages did not follow a mechanical prescription, but was based on dynamic sensing and reaction to changes in subjective scores, physiological load signals and behavioral compliance data. This organized optimization system is extremely efficient in enhancing the flexibility as well as the safety of the intervention, providing a model structure of individualized treatment that can be executed in complex rehabilitation processes.

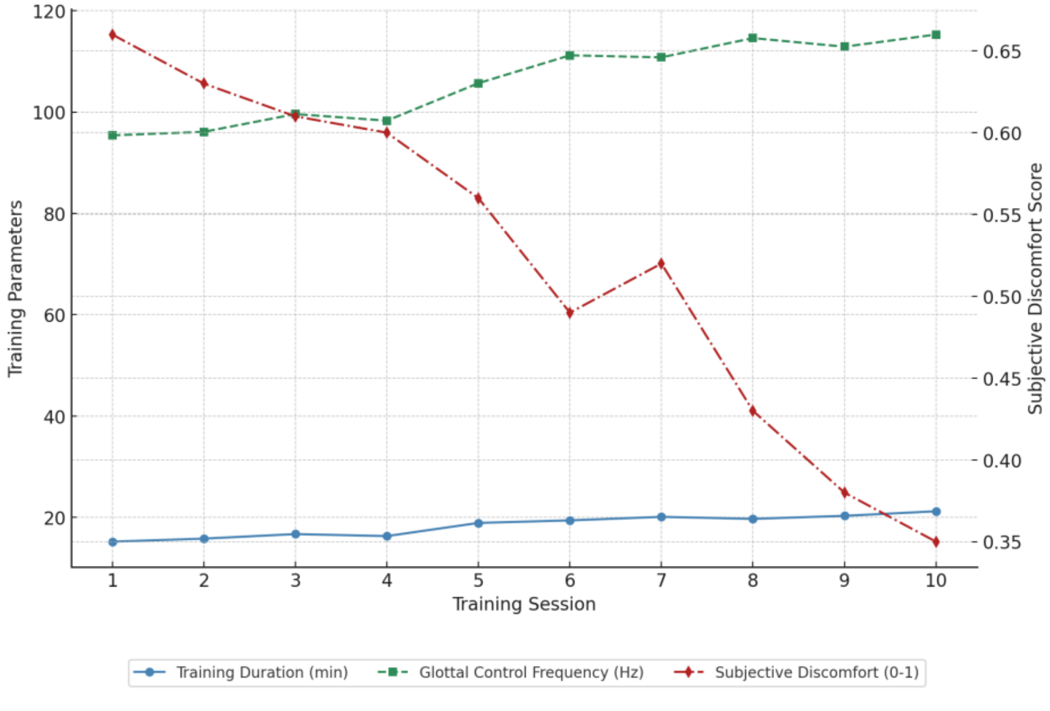

Figure 5: Dynamic evolution process of training parameters and subjective scores in personalized voice rehabilitation cases

5 Discussion

The current study suggests a multi-objective optimization model that combines an enhanced version of Non-dominated Sorting Genetic Algorithm (NSGA-VR) with acoustic, physiological and subjective perceptual aspects to create a cross-modal assessment system and to find Pareto-optimal parameters of vocal rehabilitation. The findings suggest that the quality of solutions, convergence, uniformity of distribution, and individual adaptability are improved with the NSGA-VR as compared with traditional optimization algorithms such as NSGA-II, MOEA/D, and SPEA2. Additionally, it is dynamically and personally evolving highly in real life clinical conditions. The results not only testify to the practicability and usefulness of the multi-objective optimization algorithm in the sphere of voice rehabilitation but also provide theoretical grounds and methodological support to construct intelligent and precise models of vocal rehabilitation.

The convolutional neural network (CNN) model was used by Chen et al. (2024) to predict the outcome of interventions aimed at improving vocal dysfunction using acoustic parameters as input variables and developing a model to predict the trend in rehabilitation prior to and following the intervention [26]. Their research gave theoretical basis to treatment related information extraction of the acoustic signal in this study. But their method was based on single-objective fitting prediction without considering physiological measures or subjective perception and the prediction outcomes did not add up to the formation of intervention pathways. On the other hand, this paper does not just create a tri-objective optimization model based on acoustic perturbation, physiological load, and subjective assessment, but also attains cooperative optimization of various treatment objectives in the solution space using evolutionary algorithms, which greatly enhances the personalization of pathways as well as adapts them to the needs. Behlau et al. (2023) suggested a dynamic perception-based parameter adjustment mechanism in the voice rehabilitation training, changing the training strategy through glottal image analysis and highlighting the significance of visual feedback in the vocal exercise [27]. However, their method does not have a systematic regulation platform of multi-objectives. This paper extends their work by adding the physiological signals and subjective ratings to the objective function of the model, as well as applying the non-dominated optimization search with an adapted NSGA-II algorithm, which goes beyond the rule-based changes, and into optimization-based modeling. Personalized voice therapy is one of the necessities identified by Wang et al. (2023), who noted that multifactorial factors (e.g., respiratory-glottal coordination and patient psychological compliance) are strong modulators of intervention effectiveness and recommended dynamic parameter intervention methods to enhance the rate of treatment response [28]. This paper is closely similar to this concept as it emphasizes on pathway flexibility to unique conditions. However, Wang et al. did not provide any specific modeling framework, mostly providing descriptive analyses of personalized interventions without algorithmic support. Conversely, this paper develops a Pareto front solution set in the objective space that incorporates various strategies of trade-off preference to achieve explicit criteria of pathway adjustment between individuals in different stages of rehabilitation. Further empirical findings indicate the presence of the model ability to fine tune training parameters over time.

The paper extends the use of limits of multi-objective optimization algorithms in the multi-disciplinary area of medical engineering and intelligent rehabilitation, developing biomedical problem solving as a monolithic efficacy-oriented paradigm into a multidimensional cooperation paradigm. Moreover, the study goes beyond the linear and experience-driven design model of conventional rehabilitation programs by incorporating nonlinear modelling and Pareto optimal solutions concepts, thus, creating an organized optimization model of individualized treatment programs. Methodologically, the given study ensures a high level of integration between intelligent algorithms and multimodal human-machine data, which gives a general methodological approach to the development of dynamic intervention strategies in complex feedback systems. Also, it enables the conversion of clinical rehabilitation pathway as a stagnant standardization into dynamic intelligence, serving as a working reference and technical standard of data-driven precision medicine systems.

Nevertheless, there are a number of constraints. To begin with, the experimental data sources are quite focused, which might limit the ability of the model to adjust to more general individual differences. Secondly, the optimization algorithm has high computational resource requirements and is complex to tune its parameters that make it difficult to deploy quickly and scale to general clinical environments. Lastly, although the evaluation index system has acoustic, physiological and subjective dimensions, clinical applicability of some of the key variables still can be compromised due to the problems of accessibility and consistency in measurements, which can influence external validity of the model.

6 Conclusions

The model of personalized vocal rehabilitation based on multi-objective optimization algorithm shows that it has great performance benefits when applied in practice and meets the objective of multi-dimensional collaborative optimization in therapy pathways. The key results are summarized below:

To begin with, the NSGA-VR algorithm proves its superiority in terms of multidimensional evaluation metrics over benchmark algorithms. Spread metric NSGA-VR also gets the best value (0.212), which means that solutions are well-distributed without any form of spatial clustering bias. In terms of prediction accuracy, it is much higher than other models (MSE=0.008) which verifies the improved ability of the model to observe nonlinear associations between acoustic and physiological data. The rate of ΔTCT improvement is 0.273 indicating considerable increases in execution of tasks efficiency after personal rehabilitation courses. The ΔEMG measurement value (18.362uV) is indicative of a high level of positive change in vocal fold muscle relaxation and naturalness of vocalization.

Secondly, the model produces extensive and strategically different Pareto-optimal solution sets allowing dynamic adjustment of trade-offs between acoustic perturbation, physiological load, and subjective perception depending on individualized priorities in therapy. The case evolution outcomes demonstrate that the model has a capability to perform stepwise parameter optimization, and it reacts to the changing physiological states by adjusting the intensity and frequency of training in real time. The personalized pathway provides a progressive rehabilitation process in three developmental phases: "compliance-acquisition→ adaptation → active engagement." Such an approach can be highly effective in terms of therapeutic improvement, which is indicated by the decrease in the subjective discomfort indicator values, i.e., 0.66 to 0.35.

This paper demonstrates the efficacy and versatility of multi-objective optimization algorithms in creating personalized treatment schemes to treat voice disorders, providing a theoretical foundation of smart rehabilitation solutions. The further research can follow these directions:First, the involvement of patients of different regions with different professional singing experience and pathological conditions would help to make the model more applicable and resilient in its application to heterogeneous groups.Second, the use of lightweight modeling methods and parallel processor systems could assist in simplifying the construction of the optimization algorithm and its computational architecture, which could prove more effective in execution on typical medical computers and simpler to implement at primary care sites.Finally, it has been proposed that the clinical applicability of the assessment system ought to be enhanced. Generalized approaches to measurement of physiological parameters or data alternatives on portable devices will enhance the consistency of the main indicator measurements in different environments, and increase the external validity of the model.

References

[1] Levett, J., & Pring, T. (2023). Amateur choir singers–Does good vocal health matter?. International Journal of Language & Communication Disorders, 58(4), 1286-1293. https://doi.org/10.1111/1460-6984.12863.

[2] Cazden, J. (2022). Revising Authenticity: Case Studies of Media Actors in Voice Therapy. Voice and Speech Review, 17, 151 - 166. https://doi.org/10.1080/23268263.2022.2091708.

[3] Balaji, N., Nithish, P., & Yadav, G. N. (2024, July). Anomaly Detection in Voice Conversations Through Spectral Features. In 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) (pp. 124-131). IEEE. https://doi.org/10.1109/ICSCSS60660.2024.10625647.

[4] Paulino, J. A. S., Gomes, H. M., Batista, L. V., & Lopes, L. W. (2024, September). Analysis of Frequency Range Effect on the Detection of Voice Disorder Using Convolutional Neural Networks Trained on Spectogram Images. In 2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 01-06). IEEE. https://doi.org/10.1109/SIBGRAPI62404.2024.10716317.

[5] Knuijt, S., Nonnekes, J., van de Warrenburg, B. P. C., Nijkamp, M., Scholten, S., de Swart, B. J. M., & Kalf, J. G. (2024). The use of the voice trainer app for vocal control in people with a degenerative ataxia: a pilot intervention study. The Cerebellum, 23(4), 1490-1497. https://doi.org/10.1007/s12311-024-01662-5.

[6] Wolfberg, J., Whyte, J., Doyle, P., Gherson, S., Muise, J., Petty, B., ... & Van Stan, J. H. (2024). Rehabilitation treatment specification system for voice therapy: application to everyday clinical care. American Journal of Speech-Language Pathology, 33(2), 814-830. https://doi.org/10.1044/2023_AJSLP-23-00283.

[7] Liu, G. S., Jovanovic, N., Sung, C. K., & Doyle, P. C. (2024a). A scoping review of artificial intelligence detection of voice pathology: challenges and opportunities. Otolaryngology–Head and Neck Surgery, 171(3), 658-666. https://doi.org/10.1002/ohn.809.

[8] Calà, F., Frassineti, L., Cantarella, G., Buccichini, G., Battilocchi, L., Manfredi, C., & Lanatà, A. (2025). Towards an explainable Artificial intelligence system for voice pathology identification and post-treatment characterisation. Biomedical Signal Processing and Control, 104, 107530. https://doi.org/10.1016/j.bspc.2025.107530.

[9] Calado, M., Marques, C., Pombo, J., Faria, J., & Mariano, S. (2024). A Novel

Multi-Swarm and Multi-Objective Optimization Algorithm for Solving Engineering Problems. 2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 1-7. https://doi.org/10.1109/EEEIC/ICPSEurope61470.2024.10751277.

[10] Hou, B., Zhang, Y., Lu, Y., Jie, J., Zheng, H., & Zhang, M. (2024). Gaussian Adaptive Strategy Based Multi-Objective Evolutionary Optimization for Path Planning on Uneven Terrains. IEEE Robotics and Automation Letters, 9, 539-546. https://doi.org/10.1109/LRA.2023.3334675.

[11] Ni, H., Zhou, Y., Wang, W., Wang, Z., Wu, T., & Shen, Y. (2024). Application of enhanced recovery after surgery based on multiphase optimization strategy in the nursing management of children with snoring disease day surgery. International journal of pediatric otorhinolaryngology, 186, 112123 . https://doi.org/10.1016/j.ijporl.2024.112123.

[12] Yu, X., Meng, W., Liu, Z., Ai, Q., & Liu, Q. (2021). Multi-objective Trajectory Optimization of Redundant Manipulator for Patient Assistance. 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 66-71. https://doi.org/10.1109/M2VIP49856.2021.9665073.

[13] Abed, B. M., & Jasim, W. M. (2023). Hybrid approach for multi-objective optimization path planning with moving target. Indonesian Journal of Electrical Engineering and Computer Science, 29(1), 348-357. https://doi.org/10.11591/ijeecs.v29.i1.pp348-357.

[14] Fan, Y., Zhu, L., Wang, H., & Song, A. (2023). Individualized Generation of Upper Limb Training for Robot-assisted Rehabilitation using Multi-objective Optimization. 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 921-922. https://doi.org/10.1109/VRW58643.2023.00303.

[15] Nickelson, A., Smart, W., & Tumer, K. (2023). Contextual Multi-Objective Path Planning. 2023 IEEE International Conference on Robotics and Automation (ICRA), 10240-10246. https://doi.org/10.1109/ICRA48891.2023.10160781.

[16] Wang, Z., Zhang, M., Liang, S., Yu, S., Zhang, C., & Du, S. (2025). A Multi-Objective Path-Planning Approach for Multi-Scenario Urban Mobility Needs. Algorithms, 18(1), 41. https://doi.org/10.3390/a18010041.

[17] Wang, J., & Xu, M. (2022). Learning strategy for continuous robot visual control: A multi-objective perspective. Knowl. Based Syst., 252, 109448. https://doi.org/10.1016/j.knosys.2022.109448.

[18] Widanage, K., Sheng, Z., Chen, H., & Li, Y. (2023). Multi-Objective Optimization-Based Assist-as-Needed Controller for Improved Quality of Assistance in Rehabilitation Robotics. 2023 International Conference on Rehabilitation Robotics (ICORR), 1-6. https://doi.org/10.1109/ICORR58425.2023.10304734.

[19] Li, S., Huang, S., Huang, L., Shen, H., Liu, Y., & Xie, L. (2023). Design and optimization of a body weight support system for lower-limb rehabilitation robots

considering vibration characteristics. Structural and multidisciplinary optimization, 66(12), 249. https://doi.org/10.1007/s00158-023-03700-y.

[20] Chen, Z., Zhu, P., Qiu, W., Guo, J., & Li, Y. (2023). Deep learning in automatic detection of dysphonia: Comparing acoustic features and developing a generalizable framework. International Journal of Language & Communication Disorders, 58(2), 279-294. https://doi.org/10.1111/1460-6984.12783.

[21] Matos, K. C., Oliveira, V. F. D., Oliveira, P. L. C. D., & Braga Neto, P. (2022). An overview of dysphagia rehabilitation for stroke patients. Arquivos de Neuro-Psiquiatria, 80, 84 - 96. https://doi.org/10.1590/0004-282X-ANP-2021-0073.

[22] Chen, B., Frost, N., Suolang, D., & Faigle, R. (2022). Practice Patterns and Attitudes Among Speech–Language Pathologists Treating Stroke Patients with Dysphagia: A Nationwide Survey. Dysphagia, 37, 1715 - 1722. https://doi.org/10.1007/s00455-022-10432-6.

[23] Zhang, J. (2023). The Impact of Speech Therapy Combined with Audiovisual Integration Rehabilitation Training on the Intelligence and Language Recovery of Children with Global Developmental Delay. Journal of Clinical Medicine Research, 4(2), 37-40. https://doi.org/10.32629/jcmr.v4i2.1233.

[24] Liu, Y., Hou, R., Yu, Q., & Niu, Q. (2024b). Effect of voice training intervention on swallowing function in patients with head and neck cancer undergoing radiotherapy: A randomized controlled trial. European journal of oncology nursing: the official journal of European Oncology Nursing Society, 70, 102551. https://doi.org/10.1016/j.ejon.2024.102551.

[25] Srinivasan, Y., Liu, A., & Rameau, A. (2024). Machine learning in the evaluation of voice and swallowing in the head and neck cancer patient. Current Opinion in Otolaryngology & Head and Neck Surgery, 32(2), 105-112. https://doi.org/10.1097/MOO.0000000000000948.

[26] Chen, L., Liang, Y., Dui, H., & Yang, Z. (2024). Reliability analysis and preventive maintenance of rehabilitation robots. Reliab. Eng. Syst. Saf., 256, 110704. https://doi.org/10.1016/j.ress.2024.110704.

[27] Behlau, M., Madazio, G., & Yamasaki, R. (2023, September). Dynamic vocal analysis: vocal functionality evaluation. In CoDAS (Vol. 35, p. e20210083). Sociedade Brasileira de Fonoaudiologia. https://doi.org/10.1590/2317-1782/20232021083pt.

[28] Wang, X., Cheng, C., Liu, D., Chen, Y., & Guan, X. (2023). Preliminary analysis of the effect of individualized voice therapy on pediatric voice disorders. Lin Chuang er bi yan hou tou Jing wai ke za zhi= Journal of Clinical Otorhinolaryngology, Head, and Neck Surgery, 37(5), 350-353. https://doi.org/10.13201/j.issn.2096-7993.2023.05.006.