Computational Energy Science

Vol. 2, No. 3, p. 98-110, 2025

Original article

A hybrid PSO-LSSVM framework for enhanced intelligent recognition of Vapor-Liquid Two-Phase flow patterns in geothermal production wellbores

Shengpeng He[1], Jianguo Li[2], Runjing Guan[1], Zengli Wang[1], Liang Gong[1], Tao Zhang[1]*, Shuyu Sun[3]*

[1]*College of New Energy, China University of Petroleum (East China), Qingdao 266580, P. R. China*

[2]*Qingdao Haier Refrigerator Co., Ltd., Qingdao 266580, P. R. China*

[3]*School of Mathematical Sciences, Tongji University, Shanghai 200092, P. R. China*

Keywords:
Flow patterns
geothermal production wellbore
PSO
LSSVM

Cited as:
He, S., Li, J., Guan, R., Wang, Z., Gong, L., Zhang, T., Sun, S. A hybrid PSO-LSSVM framework for enhanced intelligent recognition of Vapor-Liquid Two-Phase flow patterns in geothermal production wellbores. Computational Energy Science, 2025, 2(3): 98-110.
https://doi.org/10.46690/compes.2025.03.03

Abstract:
Accurate flash evaporation flow pattern recognition is critical for optimizing industrial processes in geothermal energy and chemical engineering. This study integrates signal processing with machine learning to advance this task. First, complementary ensemble empirical mode decomposition (CEEMD) was applied to extract intrinsic mode function (IMF) energy spectra from differential pressure signals, capturing dynamic phase-change features. Subsequent analysis revealed sample type and size significantly impact model performance: Multi-parameter datasets (combining inlet temperature, flow velocity, and IMF energy) yielded optimal training, while increased sample volume improved accuracy across all models. Three machine learning models, support vector machine (SVM), least squares SVM (LSSVM), and particle swarm-optimized LSSVM (PSO-LSSVM), were compared. SVM showed high sensitivity to sample quality; LSSVM enhanced stability but remained limited. PSO-LSSVM, however, outperformed both by leveraging PSO to optimize hyperparameters, achieving 97% test accuracy with robust generalization. It effectively distinguished flow patterns (single-phase, bubble, slug-churn, annular) using multi-feature inputs, even for complex regimes. This work provides a reliable technical framework for real-time flash evaporation monitoring and severity assessment in production wells, bridging lab insights with industrial application.

1. Introduction

Enhanced Geothermal Systems (EGS) (Dong et al., 2025; Liu et al., 2025) have emerged as a critical renewable energy technology to address global energy transition demands, as they enable the extraction of thermal energy from deep subsurface formations for power generation and heating applications. A key challenge in EGS operation lies in the dynamic behavior of working fluids within production wells: High-temperature, high-pressure geothermal fluids often undergo rapid flash evaporation (flash boiling) (Wang et al., 2021; Singh et al., 2024) in the near-wellbore section due to sudden pressure drops as fluids flow from the high-pressure reservoir to the wellbore (Wang, 2023). These flow patterns are not merely structural manifestations of two-phase flow but direct indicators of the intensity and stability of the flash evaporation process-making their accurate recognition pivotal to EGS operational safety and efficiency (Liu et al., 2025). The stability of vapor-liquid two-phase flow patterns in geothermal production wells exerts a profound impact on both geothermal extraction efficiency and the long-term integrity of the production system (Singh et al., 2025). For instance, unsteady flow patterns (such as slug flow or churn flow (Elgaddafi et al., 2025)) can induce periodic pressure oscillations within the wellbore (Marcel et al., 2010), leading to increased mechanical stress on well

Yandy Scientific Press

*Corresponding author.
E-mail address: B25150012@s.upc.edu.cn (S. He); 1229015617@qq.com (J. Li); by2115020104@s.upc.edu.cn (R. Guan); wangzengli25@upc.edu.cn (Z. Wang); lgong@upc.edu.cn (L. Gong); tao.zhang@upc.edu.cn (T. Zhang).

Received July 10, 2025; revised August 7, 2025; accepted September 8, 2025; available online September 13, 2025.

casing and downhole equipment, accelerated corrosion, and even catastrophic failures such as pipe rupture. Additionally, irregular flow patterns disrupt the uniform distribution of thermal energy in the fluid, reducing the heat exchange efficiency in surface power cycles and thus lowering the overall energy output of the EGS. Conversely, stable flow patterns (e.g., annular flow under controlled conditions) facilitate consistent fluid transport and heat transfer, optimizing the energy conversion process. Given these consequences, investigating the recognition and classification of vapor-liquid two-phase flow patterns induced by forced circulation flash evaporation is not only a matter of academic interest but also a practical necessity for ensuring the reliability, safety, and economic viability of geothermal production systems.

For decades, the study of vapor-liquid two-phase flow patterns has relied on two primary methodological frameworks: traditional experimental observation and numerical simulation (Ma et al., 2025). However, both approaches exhibit significant limitations when applied to the complex, extreme conditions of geothermal production wells, particularly in the context of flash evaporation-induced flow patterns. Traditional experimental methods for flow pattern recognition can be broadly categorized into direct observation and indirect measurement techniques, with direct observation being the earliest and most intuitive approach. Direct observation methods typically involve visual inspection or high-speed camera imaging (Tong and Zhu, 2025), which capture flow patterns by leveraging transparent test sections in laboratory setups. While these methods offer high spatial and temporal resolution for flow pattern visualization under controlled conditions, their applicability to geothermal production wells is severely constrained. First, geothermal wellbores are constructed from non-transparent, high-strength metallic materials to withstand high temperatures and pressures, eliminating the possibility of direct visual access to internal flow. Second, the high flow velocities and violent flash evaporation processes in geothermal wells result in highly dynamic, transient flow patterns that exceed the temporal resolution limits of conventional high-speed cameras, which are typically optimized for steady or slowly varying flow in laboratory environments. Third, the extreme downhole conditions (high temperature, high pressure, and corrosive fluids) pose significant challenges to the durability and reliability of observation equipment; even specialized high-temperature cameras would degrade rapidly under such conditions, leading to inconsistent or invalid data. These limitations render direct observation methods impractical for real-world geothermal well applications, restricting their utility to laboratory-scale validation rather than in-situ monitoring.

Numerical simulation methods, particularly computational fluid dynamics (CFD) models, have emerged as an alternative to experimental observation, aiming to replicate two-phase flow behavior through mathematical modeling of fluid dynamics, heat transfer, and phase transition. Common CFD approaches for two-phase flow include the volume of fluid (VOF) method (Yuan et al., 2025), Eulerian-Eulerian two-fluid model (Yan et al., 2025), and Lagrangian particle tracking model (Caraghiaur and Anglart, 2013), which differ in their treatment of phase interfaces and computational complexity. While these models have demonstrated success in simulating steady-state two-phase flow patterns (e.g., bubbly flow in low-pressure pipelines) under laboratory conditions, they face substantial challenges when applied to flash evaporation in geothermal production wells. First, flash evaporation is a highly transient, thermodynamically non-equilibrium process, where the rapid pressure drop causes instantaneous phase change-requiring precise modeling of heat and mass transfer across phase interfaces. Most existing CFD models simplify phase transition kinetics by assuming local thermodynamic equilibrium, which fails to capture the rapidity and non-uniformity of flash evaporation in geothermal wells, leading to inaccurate predictions of flow pattern formation and evolution. Second, the computational cost of high-fidelity two-phase flow simulations is prohibitive for large-scale geothermal wellbore models. Third, CFD models rely heavily on empirical closure relations (e.g., drag coefficients, interfacial heat transfer correlations) calibrated from laboratory experiments, which are often derived under conditions (e.g., low temperature, low pressure, air-water systems) that differ drastically from the high-temperature, high-pressure geothermal fluid systems (e.g., water-steam or brine-steam mixtures). This mismatch in operating conditions leads to significant uncertainties in model predictions, reducing the reliability of CFD-based flow pattern recognition for geothermal wells. In summary, traditional experimental methods suffer from inherent limitations in adapting to the non-transparent, extreme conditions of geothermal production wells, while numerical simulation methods are constrained by computational cost, simplifying assumptions, and the lack of validated closure relations for flash evaporation processes. These gaps highlight the need for alternative, more robust methods that can overcome the limitations of traditional approaches and provide accurate, in-situ flow pattern recognition for geothermal production wells.

In response to the limitations of traditional experimental and numerical methods, indirect measurement technique, particularly those based on differential pressure fluctuation signals have emerged as a promising alternative for two-phase flow pattern recognition (Zaborowska et al., 2020; Camargo and Paladino, 2025). Differential pressure signal-based methods leverage the principle that different flow patterns generate distinct pressure fluctuation characteristic. By installing pressure sensors at two points along the flow path to capture these fluctuation signals, researchers can extract flow pattern-specific features and establish correlations between signal characteristics and flow patterns-offering a practical, cost-effective, and non-intrusive solution for in-situ monitoring, especially in non-transparent pipelines like geothermal wellbores (Liu et al., 2025). Over the past few decades, this approach has undergone significant advancements, driven by progress in signal processing techniques and the integration of machine learning, while also facing persistent challenges related to adaptability, robustness, and applicability to geothermal-specific conditions. These advancements have transformed differential pressure signal-based methods from manual, semi-quantitative techniques to intelligent, data-driven tools, with clear trends toward multi-feature fusion, deep learning for automatic feature extraction, and adaptability to complex

phase transition processes like flash evaporation. Despite these advancements, several critical challenges remain, particularly in adapting these methods to the specific conditions of geothermal production wells. First, most existing studies are conducted under laboratory conditions with simplified working fluids, which differ drastically from the high-temperature, high-pressure, and high-salinity geothermal fluidsin production wells. This mismatch leads to poor generalization of trained models, as features extracted from laboratory signals may not be representative of flash-induced flow patterns in geothermal wells. Second, existing methods primarily target steady-state boiling processes, while flash evaporation in geothermal wells is a highly transient, non-equilibrium process-current signal processing and machine learning models struggle to capture the rapid, dynamic changes in pressure signals associated with transient flash events, leading to delayed or inaccurate flow pattern classification. Third, multi-feature fusion approaches often lack a systematic framework for weight assignment, with auxiliary parameters arbitrarily combined with differential pressure signals-resulting in redundant features or information loss that degrades model performance. Fourth, the robustness of machine learning models to downhole disturbances is insufficient; most models are trained on clean, controlled laboratory data and fail to maintain accuracy under the noisy, variable conditions of geothermal wells.

Addressing these challenges is essential for advancing differential pressure signal-based methods from laboratory research to practical geothermal well applications. This study advances intelligent flash evaporation flow pattern recognition by bridging experimental insights with machine learning. We developed a forced-circulation visual platform to generate differential pressure signals under controlled flash conditions, then applied CEEMD to extract IMF energy spectra, resolving phase-dynamics coupling in two-phase flow. Through systematic sample design and model comparison (SVM, LSSVM, PSO-LSSVM), we demonstrated PSO-LSSVM's superiority (Gorjaei et al., 2015), outperforming others by mitigating overfitting and enhancing generalization. Key contributions include: Validating CEEMD for feature extraction, defining optimal multi-parameter datasets, and proving PSO-LSSVM's robustness for flow pattern distinction. These outcomes lay groundwork for addressing future challenges, field-condition adaptation, transient signal processing, and multimodal fusion, ultimately positioning pressure-signal-based methods as reliable tools for geothermal well monitoring.

2. PSO-LSSVM framework with CEEMD algorithm

To realize rapid and accurate prediction of in-pipe flow pattern transitions, a particle swarm optimization-least squares support vector machine (PSO-LSSVM) framework integrated with complementary ensemble empirical mode decomposition (CEEMD) for feature engineering is developed. The technical process is articulated as follows.

2.1 Flow pattern feature database construction

Differential pressure fluctuation signals from wellbore flash evaporation-induced flow patterns are inherently nonlinear and non-stationary, posing challenges for traditional processing techniques: Fourier Transform fails due to its stationarity assumption, while Wavelet Decomposition suffers from subjective basis function selection, both unable to extract valid flow pattern features. To address this, Huang et al. (1998) proposed EMD, a data-driven, adaptive method that decomposes signals into IMFs (capturing distinct time-frequency oscillatory modes) and a residual, preserving fine-grained details critical for flash flow pattern recognition. However, EMD suffers from "mode mixing" (single IMF contains multiple modes) when processing flash signals: Sudden pressure spikes (e.g., slug passage) and sensor noise disrupt its sifting process, corrupting IMF features and leading to flow pattern misclassification. Wu and Huang (2009) developed EEMD to mitigate mode mixing by adding small-amplitude white noise before EMD, smoothing signal discontinuities. Yet EEMD has drawbacks for flash data: Hundreds of iterations drastically increase computation (incompatible with real-time wellbore monitoring) and residual white noise remains in low-signal-to-noise flash signals, distorting features.

CEEMD resolves EEMD's limitations by adding pairs of positive/negative white noise. This complementary strategy cancels noise without massive iterations, reducing computation and eliminating residual noise-perfectly adapting to wellbore flash signals' demands for real-time processing and high-fidelity feature extraction. The detailed process of CEEMD can be described as follows.

Step 1: Adding the white noise data series $n_i(t)$ into the original data $x(t)$, to construct the new signal $s_i(t)$ (or $w_i(t)$):

$$s_i(t) = x(t) + n_i(t) \tag{1}$$

$$w_i(t) = x(t) - n_i(t) \tag{2}$$

Step 2: Using the EMD approach to decompose $s_i(t)$, $w_i(t)$, resulting K IMF indexes:

$$s_i(t) = \sum_{j=1}^{k} IMF_{j,i}^{+}(t) \tag{3}$$

$$w_i(t) = \sum_{j=1}^{k} IMF_{j,i}^{-}(t) \tag{4}$$

If EMD is not converged, continue adding a different white noise into the original signal $x(t)$ and repeat Step 2 with $i = i + 1$, until the decomposing is ended when EMD is converged.

Step 4: Calculate the averaged IMF value after n EMD decompostion:

$$IMF_j = \frac{1}{2n} \sum_{i=1}^{n} \left(IMF_{j,i}^{1} + IMF_{j,i}^{2} \right) \tag{5}$$

Modal decomposition of differential pressure fluctuation signals for each flow pattern using CEEMD allows for the acquisition of decomposition plots corresponding to each IMF component of each flow pattern. However, the IMF components obtained directly at this stage cannot be directly

employed as feature quantities for flow pattern recognition. To further extract valid features, it is essential to calculate the energy of each IMF component. The energy features of IMF components exhibit high sensitivity to subtle changes in signals, enabling the capture of fine differences between distinct flow patterns. Meanwhile, these energy features demonstrate strong robustness against signal noise interference and end effects, which reduces the impact of such interferences to a certain extent. Specifically, the energy of an IMF component can be computed as the sum of the squares of the amplitudes of all sampling points within that particular IMF component:

$$E_i = \sum_{k=1}^{N} x_k^2 \tag{6}$$

2.2 LSSVM Framework

Support vector machines (SVM) are a class of machine learning algorithms centered on the principle of maximizing the classification margin, excelling in handling high-dimensional data and nonlinear relationships. Its foundational formulation, known as hard margin SVM, assumes data is linearly separable and aims to find an optimal hyperplane that separates different classes with the largest possible distance. The key to this hyperplane lies in "support vectors" — the data points closest to the hyperplane-since they alone determine the hyperplane's position and margin width. However, hard margin SVM fails in real-world scenarios where data often contains noise or outliers, as it cannot tolerate any misclassification. To address this, soft margin SVM was developed: It introduces slack variables to allow minor deviations of data points from the ideal separation and incorporates a penalty parameter to balance margin maximization and misclassification minimization, enhancing robustness against noise. For cases where data is nonlinearly separable even with slack variables, nonlinear SVM provides a solution. Its core insight is mapping low-dimensional, nonlinearly separable data into a higher-dimensional feature space where linear separation becomes feasible. This mapping is achieved implicitly via kernel functions, which compute the similarity between data points in the high-dimensional space without explicit transformation-avoiding excessive computational complexity. Common kernel functions (e.g., Gaussian radial basis function) effectively capture complex nonlinear patterns, enabling SVM to generalize well to diverse tasks, such as identifying flow regimes in geothermal wells using differential pressure signals. Overall, SVM evolves from hard margin (linear separability) to soft margin (noise tolerance) and nonlinear variants (complex pattern handling), forming a flexible framework for classification.

Least squares support vector machines (LSSVM) are an optimized extension of traditional SVM, designed to simplify computational complexity while retaining SVM's strengths in nonlinear and high-dimensional data processing. The core innovation of LSSVM lies in reformulating the optimization problem of traditional SVM. In traditional soft margin SVM, the optimization involves solving a quadratic programming problem with inequality constraints, which is computationally intensive, especially for large datasets. LSSVM addresses this by replacing the inequality constraints with equality constraints and modifying the loss function: Instead of using the hinge loss (which measures misclassification errors via slack variables), it adopts a least squares loss function that squares the error terms. This transformation converts the complex quadratic programming problem into a system of linear equations, significantly reducing computational burden and accelerating model training. LSSVM also streamlines parameter solution by leveraging the Karush-Kuhn-Tucker (KKT) conditions. Unlike traditional SVM, which requires solving for Lagrange multipliers under strict inequality constraints, LSSVM derives a set of linear equations that can be solved directly-improving numerical stability and further simplifying computation. Crucially, LSSVM retains the kernel function mechanism of traditional SVM, ensuring it still effectively models nonlinear relationships between features (e.g., linking IMF energy features of differential pressure signals to flash flow patterns). By balancing performance and efficiency, LSSVM overcomes the computational limitations of traditional SVM, making it more suitable for real-time applications like in-situ flow regime recognition in geothermal production wells.

The regression of LSSVM can be modeled as the following optimization problem with the loss function ξ:

$$\begin{aligned} &\min_{\omega,b,\xi_i} \frac{1}{2}\|\omega\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 \\ &\text{s.t.} \quad y_i \cdot (\omega \cdot \phi(x_i) + b) = 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, 2, \cdots, n \end{aligned} \tag{7}$$

Integrating the Lagrange operator α_i to construct the unconstrained functions:

$$\begin{aligned} L(\omega, b, \xi, \alpha) = &\frac{1}{2}\|\omega\|^2 + \frac{1}{2}\sum_{i=1}^{n}\xi_i^2 - \\ &\sum_{i=1}^{n}\alpha_i\{y_i[\omega^T\phi(x) + b] - 1 + \xi_i\} \end{aligned} \tag{8}$$

Considering the KKT condition, calculating the partial differential to ω, b, ξ_i, α_i and let the differential to be 0:

$$\begin{cases} \dfrac{\partial L(\omega, b, \xi, \alpha_i)}{\partial \omega} = 0 \Rightarrow \omega = \displaystyle\sum_{i=1}^{n} a_i\phi(x_i) \\ \dfrac{\partial L(\omega, b, \xi, \alpha_i)}{\partial b} = 0 \Rightarrow \displaystyle\sum_{i=1}^{n} a_i = 0 \\ \dfrac{\partial L(\omega, b, \xi, \alpha_i)}{\partial \xi} = 0 \Rightarrow a_i = C\xi_i \\ \dfrac{\partial L(\omega, b, \xi, \alpha_i)}{\partial \alpha_i} = 0 \Rightarrow \omega^T\phi(x) + b + \xi_i - y_i = 0 \end{cases} \tag{9}$$

where $i = 1, 2, \cdots, n$. Eliminate the corresponding parameters ω and ξ_i, the following matrix function can be written as:

$$\begin{bmatrix} 0 & \mathbf{y}^T \\ y & \mathbf{Z}\mathbf{Z}^T + C^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{10}$$

where $y = [y_1, y_2, \cdots, y_n]$, $\mathbf{1}_N = [1, 1, \cdots, 1]^T$, $\mathbf{Z} = [\phi(x_1), \phi(x_2), \cdots, \phi(x_n)]^T$, $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T$, $\mathbf{I}$ is a $n \times n$ matrix.

As a result, the LSSVM decision function can be formulated as:

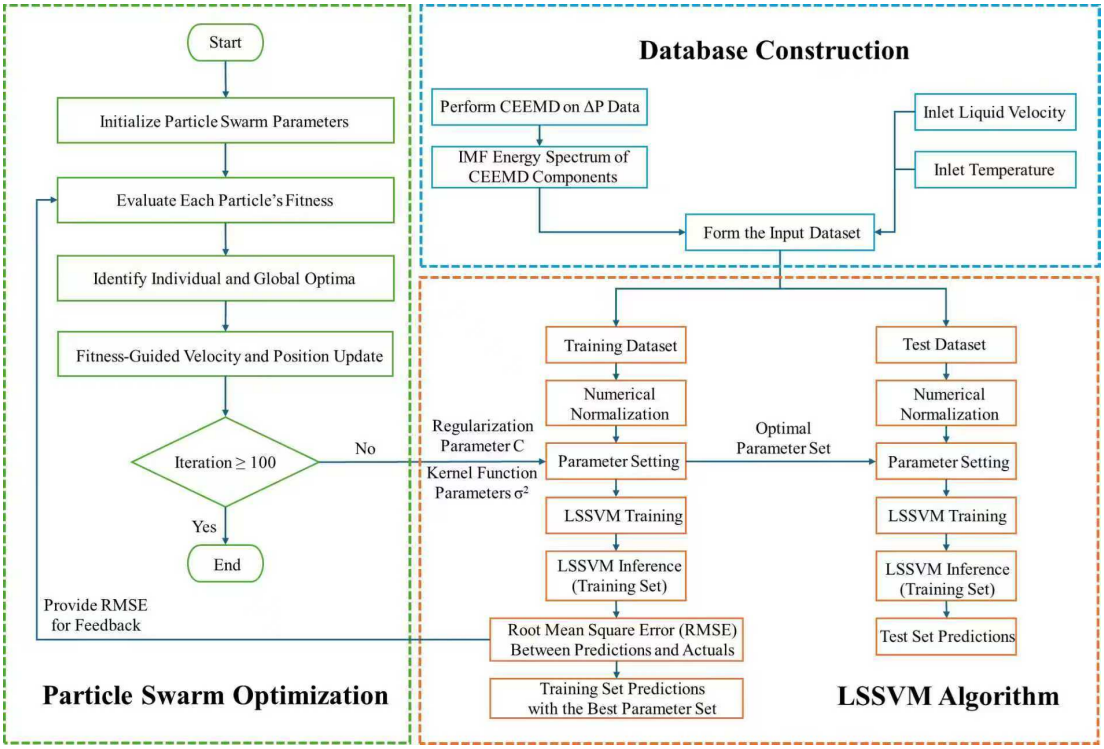


Fig. 1. Flow chart of vapour-liquid two-phase flow pattern recognition based on PSO-LSSVM using CEEMD algorithm.

$$f(x) = \text{sign}\left[\sum_{i=1}^{n} a_i y_i \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) + b\right] \quad (10)$$

2.3 PSO-based optimization of LSSVM hyperparameters

The LSSVM framework addresses two key challenges in classification tasks: Noise interference and nonlinear data relationships, by introducing a regularization parameter C and a Radial Basis Function (RBF) kernel parameter σ. Specifically, C controls the trade-off between maximizing the classification margin and minimizing noise-induced misclassifications, while the RBF kernel maps nonlinear data into a high-dimensional space where linear separation becomes feasible. However, the performance of the LSSVM model depends heavily on the specific values of C and σ. Traditional parameter tuning methods rely on manual, step-by-step adjustment and verification, which are not only time-consuming but also prone to missing the global optimal parameter combination, especially when dealing with complex data. To overcome this limitation and further enhance the model's prediction efficiency and accuracy, the Particle Swarm Optimization (PSO) algorithm is integrated with LSSVM to form the PSO-LSSVM hybrid method, leveraging the swarm intelligence of PSO to achieve adaptive and efficient optimization of LSSVM's key parameters. The inspiration for the PSO algorithm, the core of the hybrid method, originates from the collective collaborative behavior of animals during foraging. In the PSO framework, this biological behavior is abstracted into a computational model: The "particle swarm" consists of multiple individual "particles", each representing a candidate solution to the parameter optimization problem. Each particle has two key attributes: "Position" (corresponding to the parameter combination it represents) and "velocity" (determining the direction and step size of its movement in the parameter space). The core logic of PSO lies in the collaborative learning of particles: Instead of searching for optimal parameters independently, each particle updates its position and velocity by referencing two critical pieces of information: The "historical best position" (pbest) and the "global best position" (gbest). This collective learning mechanism ensures that the swarm converges toward the global optimal solution efficiently, avoiding the local optima that often plague manual tuning or single-agent search methods. By integrating PSO with LSSVM, the hybrid method not only retains LSSVM's strengths in handling high-dimensional and nonlinear data but also solves its parameter tuning bottleneck, significantly improving the model's robustness and classification accuracy. The parameter optimization process of PSO-LSSVM follows a systematic, iterative workflow, designed to efficiently narrow down the optimal C and σ combination. The updating of particle velocity and location can be formulated as follows:

$$v_{id} = \omega v_{id} + c_1 r_1 (pbest_{id} - x_{id}) + c_2 r_2 (gbest_{id} - x_{id}) \quad (11)$$

$$x_{id} = x_{id} + v_{id} \quad (12)$$

In summary, the flow chart of vapour-liquid two-phase flow pattern recognition based on PSO-LSSVM using CEEMD algorithm is presented in Fig. 1:

3. Intelligent flow pattern recognition using CEEMD and PSO-LSSVM

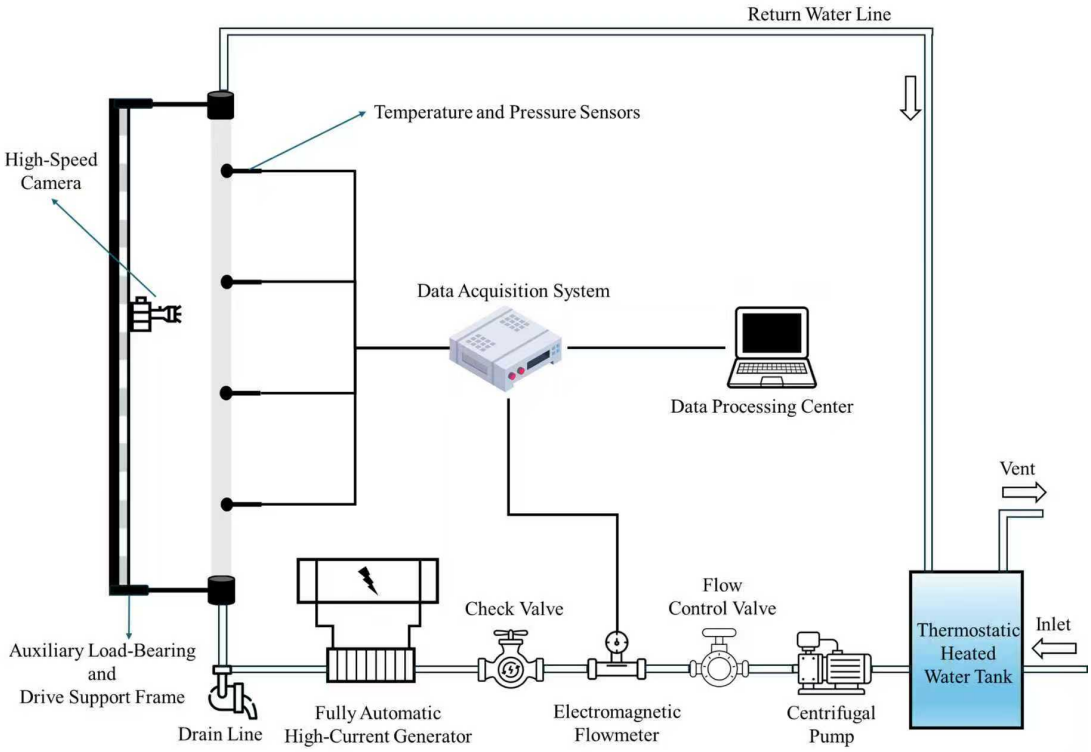


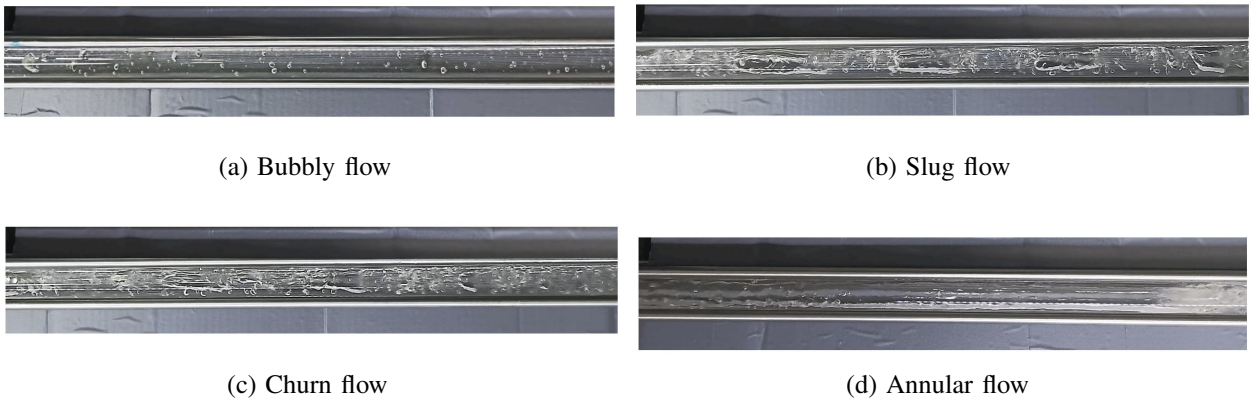
Fig. 2. Experimental system diagram.

3.1 Experiment and signal collection

To achieve controlled flow and flash evaporation of high-temperature, high-pressure fluid within the visual experimental pipeline, the entire experimental apparatus incorporates auxiliary systems including a fluid dynamic control system, a fluid temperature control system, and a dynamic data detection and acquisition system, collectively constituting a forced circulation visual flash evaporation experimental platform (as illustrated in Fig. 2). Initially, the working fluid undergoes preheating in a heating tank. Subsequently, driven by a circulation pump, the preheated fluid circulates within the system. The flow rate is precisely regulated by adjusting the opening of a proportional valve and measured accurately using an electromagnetic flowmeter. To ensure the fluid reaches the temperature required for flash evaporation, an electric heating tube performs secondary heating on the circulating fluid. Post-secondary heating, the fluid enters the visual pipeline. As the pipeline elevation increases, decreasing pressure induces flash evaporation phase change of the high-temperature fluid inside. During this phase, a data logger captures electrical signals from the electromagnetic flowmeter, pressure transducers, and thermocouples to obtain critical parameters including flow velocity, pressure, and temperature of the fluid within the pipeline. Finally, the two-phase fluid (gas-liquid coexistence) exits through the top outlet of the visual pipeline, flows back along the return pipe, and returns to the heating tank, completing a closed circulation loop.

To investigate the correlation between differential pressure and flash evaporation intensity, this study synchronously collects transient pressure data at upstream and downstream positions using pressure transducers for different flow patterns. By calculating the pressure difference between these positions, differential pressure fluctuation signals are derived, enabling analysis of their relationship with flow patterns. To expand the experimental operating range, four flow velocities (0.3 m/s, 0.4 m/s, 0.5 m/s, and 0.6 m/s, noted as condition 1, 2, 3, and 4) are achieved by adjusting the secondary heating power, generating various flash evaporation flow patterns under each condition. Differential pressure fluctuation signals for each pattern at different velocities are measured, with a sampling frequency of 150 Hz and duration of 30 s maintained to preserve the intact characteristics of the original flow pattern signals. The results are shown in Figs. 3 and 4.

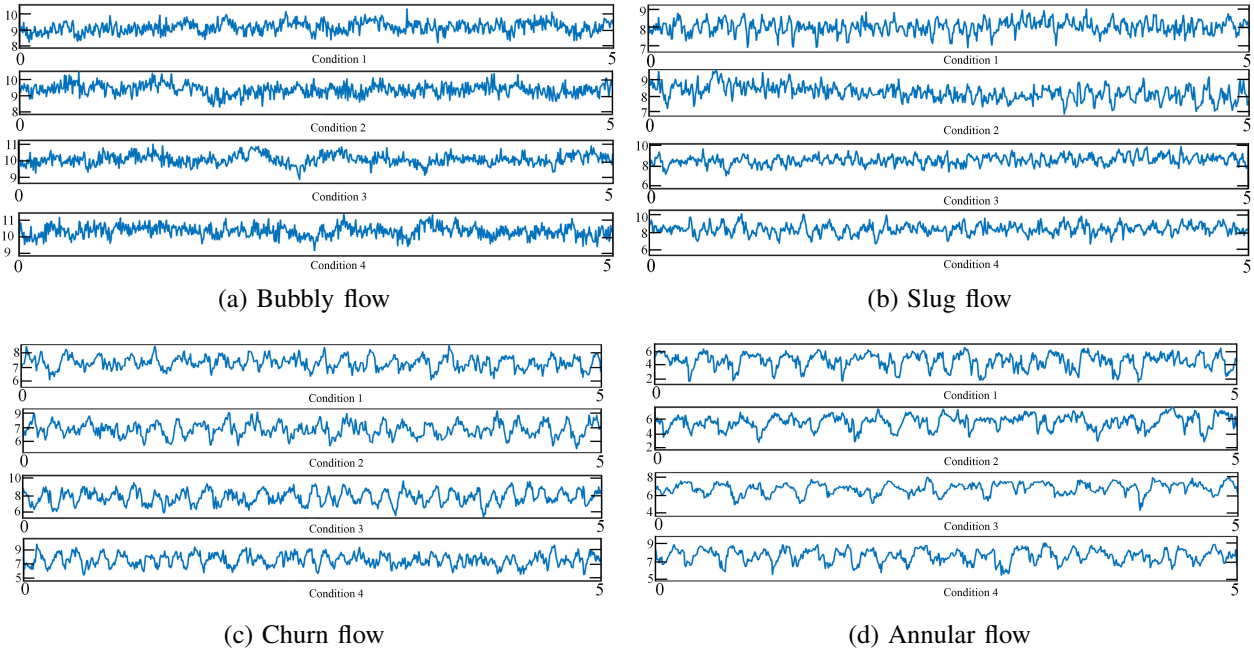
In the flash evaporation experiments of geothermal production wells' near-surface regions, differential pressure fluctuation signals exhibit distinct characteristics under varying flow velocities across different flow patterns (bubble flow, slug flow, churn flow, and annular flow), reflecting the coupling between flash evaporation phase change intensity and two-phase flow dynamics. At a low velocity of 0.3 m/s, initial bubble flow, characterized by randomly distributed small bubbles undergoing frequent breakup, generates high-frequency pressure perturbations with a differential pressure range of 8-9.5 kPa, a peak amplitude of $\sim$ 2 kPa, and the highest oscillation frequency. As flow velocity increases to 0.6 m/s, enhanced fluid kinetic energy and increased bubble quantity/volume expand the differential pressure range to 9-11 kPa (while maintaining a $\sim$ 2 kPa peak), preserving the high-frequency trait. Compared to bubble flow, slug flow, which is formed by coalescing bubbles into large "bullet-shaped" or elliptical bub-



(a) Bubbly flow (b) Slug flow

(c) Churn flow (d) Annular flow

Fig. 3. Observed results of differential pressure fluctuations.



(a) Bubbly flow (b) Slug flow

(c) Churn flow (d) Annular flow

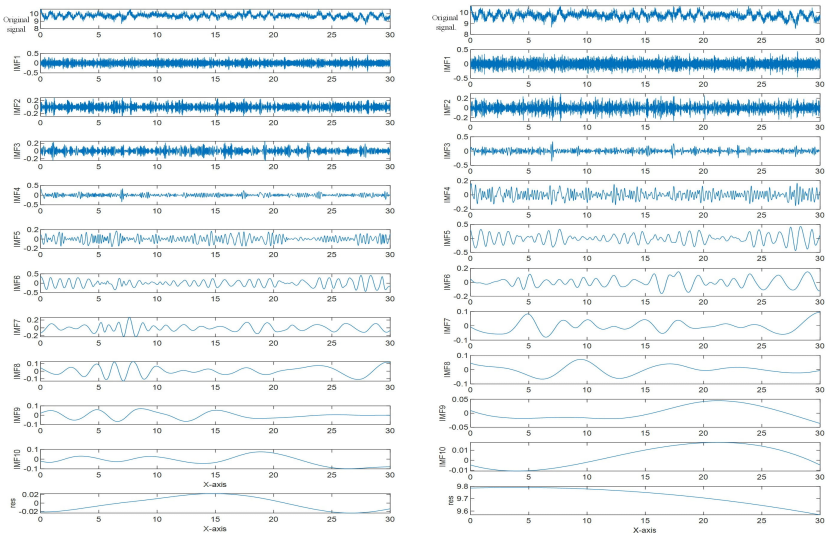
Fig. 4. Differential pressure fluctuation signal curve.

bles, exhibits significantly higher amplitude (peak increased to 3 kPa at 0.3 m/s) due to violent pressure oscillations from bubble expansion, though its lower oscillation frequency and intermittent evolution with adjacent bubble/churn flows reduce signal discriminability.

As flow transitions to churn flow, intensified turbulence from unstable liquid films causes chaotic gas-phase distribution, reducing the differential pressure range to 6-8 kPa at 0.3 m/s with a 3.5 kPa peak despite lower gravity head from increased vapor content; its disordered perturbations yield complex time-frequency structures. Finally, annular flow, which is dominated by vapor/entrained droplets in the pipe center and a stable wall liquid film that exhibits periodic pressure fluctuations from film rupture, narrowing the range to 2-6 kPa (at 0.3 m/s). However, it amplifies the peak to 4 kPa, with the lowest oscillation frequency. Cross-over phenomena further reveal limitations of absolute pressure value, for example, single-phase flow (0.3 m/s, 9.5-10 kPa) overlaps with bubble flow (0.4 m/s, 8.5-10 kPa), and slug (0.6 m/s, 9-11 kPa) partially overlaps with churn flow (6-10 kPa). Thus, accurate flash evaporation characterization requires integrating multidimensional features: Amplitude, frequency, and signal morphology, which collectively distinguish flow patterns through combined criteria rather than isolated pressure magnitudes.

3.2 Signal data processing

For the differential pressure fluctuation signal of bubble flow under an inlet velocity of 0.4 m/s and secondary heating power of 4 kW, both EMD and CEEMD methods were applied for decomposition, with results compared in Fig. 5. Visual inspection of the intrinsic mode functions (IMFs) from each decomposition reveals that CEEMD-derived IMFs display



(a) EMD (b) CEEMD

Fig. 5. Comparison of EMD and CEEMD decomposition effects.

lower vibration amplitudes and frequencies compared to EMD outputs. This indicates CEEMD more effectively suppresses mode mixing, preventing distortions from abnormally high IMF amplitudes and frequencies caused by mode aliasing. Further examination of residual terms shows EMD leaves periodic fluctuations in its residual, while CEEMD produces a residual devoid of such periodicity. This suggests EMD's decomposition remains incomplete, which may be caused by the mode mixing or endpoint effects, causing low-frequency periodic components in the residual and causing information loss. CEEMD avoids this issue, ensuring thorough signal decomposition. To reliably decompose differential pressure signals across various flow patterns and reduce environmental noise interference, CEEMD is selected for this study, mitigating mode mixing risks and enhancing decomposition accuracy.

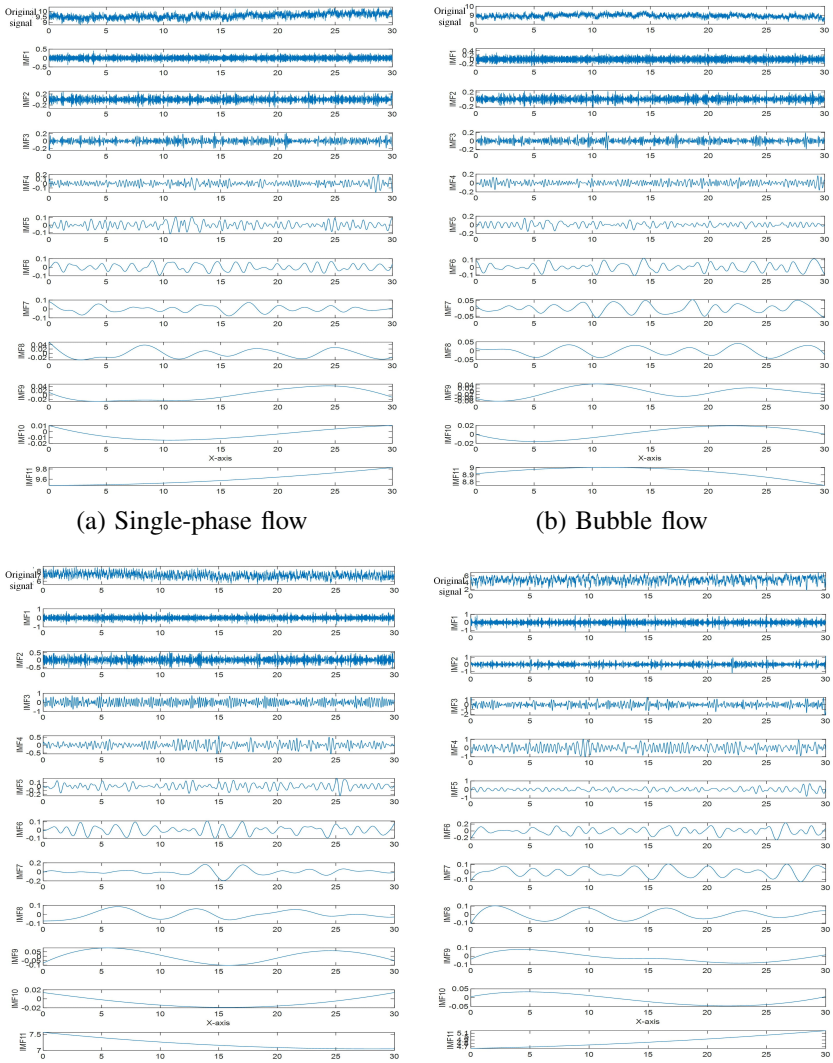
To preserve the integrity of sampled digital signals, the sampling frequency was set following the Nyquist-Shannon sampling theorem, which mandates at least twice the highest signal frequency. A sampling frequency of 150 Hz with 30-second duration was chosen to capture complete signal features. Experiments adjusted secondary heating power to maintain inlet temperatures between 95 and 115 °C at the visual pipe entrance, while flow control valves regulated liquid-phase velocity to 0.2-0.5 m/s. Flow patterns were observed along the vertical pipe, with differential pressure signals recorded between 2 m and 3 m above the pipe base for each pattern. CEEMD was used to extract IMF energy spectra from these signals as recognition features. Previous observations showed slug flow failed to stabilize as dominant within 30 seconds, intermittently transitioning with churn flow and causing significant time-frequency feature overlap. To address this, slug and churn flows were merged into a composite category, finalizing four patterns: Single-phase, bubble, slug-churn, and annular. CEEMD decomposition was then applied to each pattern's differential pressure signals, with results shown in Fig. 6.

3.3 Intelligent pattern recognition

In machine learning-based flow pattern recognition studies, recognition accuracy depends not only on sample quality but also on sample quantity. To effectively classify flash evaporation flow patterns, this section constructs a flow pattern sample library. The library includes input parameters (feature vectors of each flow pattern) and output variables (single-phase flow, bubble flow, slug-churn composite flow, and annular flow encoded as 0, 1, 2, and 3, respectively). For model training and evaluation, 80% of the data in each validation database is randomly assigned as the training set, with the remaining 20% serving as the test set. Given varying dimensions and value ranges of features, normalization is critical to prevent dominance by large-value features and ensure balanced influence of all features. Min-max normalization is applied to process features:

$$X' = \frac{X - \max\{X\}}{\max\{X\} - \min\{X\}} \tag{13}$$

To investigate how sample size affects recognition performance, three models-SVM, LSSVM, and PSO-LSSVM are trained on databases with 100, 200, 300, and 400 samples per class, and the classification results across these sample sizes are shown in Figs. 7, 8 and 9 for the three methods. The y-axis labels 0-3 represent single-phase, bubble, slug-churn composite, and annular flows, while the x-axis shows sample count. Observations indicate SVM initially misclassifies primarily single-phase and bubble flows with small samples. As sample size increases, misclassification of these two patterns decreases, but errors rise for slug-churn composite and annular flows. Overall accuracy still improves, suggesting larger samples enhance generalization by better capturing feature differences. However, SVM's limitations in handling multi-dimensional features become apparent with more data, particularly reducing its ability to recognize slug-churn com-



(a) Single-phase flow (b) Bubble flow

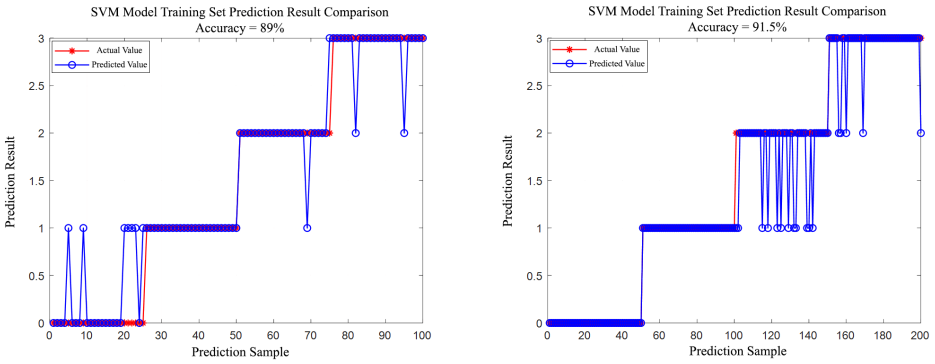
(c) Slug-churn flow (d) Annular flow

Fig. 6. CEEMD decomposition diagram of pressure difference fluctuation signals for various flow patterns.

posite and annular flows, leading to higher misclassification rates for these types. Meanwhile, as sample size increased, LSSVM's recognition accuracy improved notably. At 400 samples per class, LSSVM achieved a high accuracy of 95.5%. This gain stems from LSSVM's mechanism: By introducing error squared terms and equality constraints, it transforms SVM's quadratic programming problem into solving linear equations, streamlining computation while retaining strong generalization. For flow pattern recognition—particularly with large datasets-this simplification not only maintains accuracy but also boosts efficiency, making LSSVM more practical than SVM for handling growing experimental data.

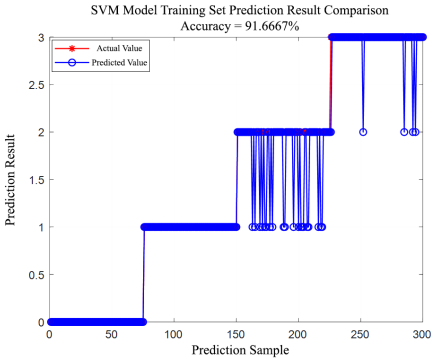
Further optimization with particle swarm optimization elevated LSSVM's performance by roughly 5% across all training sets. PSO enhances LSSVM by globally searching for optimal hyperparameters that minimize classification errors. Unlike grid search or random search, PSO's population-based iterative approach efficiently navigates parameter space, avoiding local optima and fine-tuning LSSVM's decision boundaries to better capture subtle feature differences between flow patterns. This is critical for flash evaporation scenarios, where flow patterns often exhibit overlapping characteristics in high-dimensional feature spaces. The suitability of PSO-LSSVM for phase change-related flow pattern recognition lies in its dual strengths: first, LSSVM's linear-solving efficiency handles the computational load of multi-feature, large-sample data typical in phase change experiments; second, PSO's optimization refines the model's ability to distinguish nonlinear relationships, which is the key for identifying transient or subtly evolving flow patterns during flash evaporation. Where traditional models struggle with parameter sensitivity or local optima, PSO-LSSVM's adaptive parameter tuning ensures robust performance, making it uniquely equipped to model the complex interplay of heat transfer, fluid dynamics, and phase change dynamics that define these flow patterns.

Fig. 10 presents the fitness curve of the PSO-LSSVM model during flow pattern recognition classification. The particle swarm optimization algorithm, guided by fitness feedback,

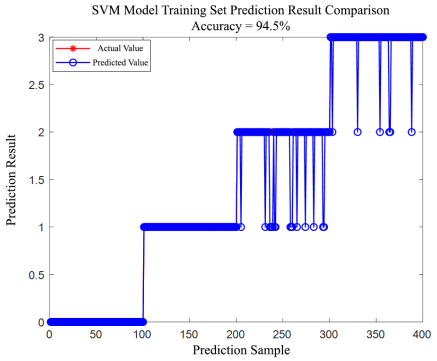


(a) 100 data

(b) 200 data

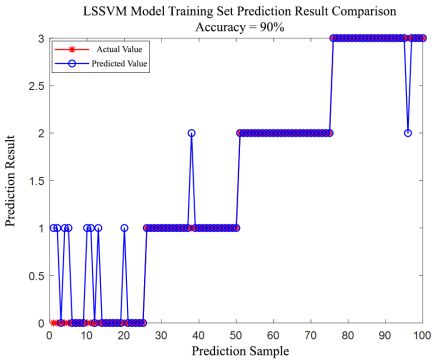


(c) 300 data

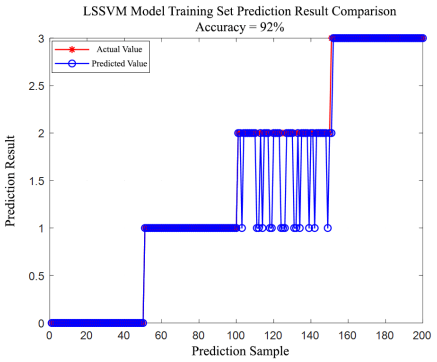


(d) 400 data

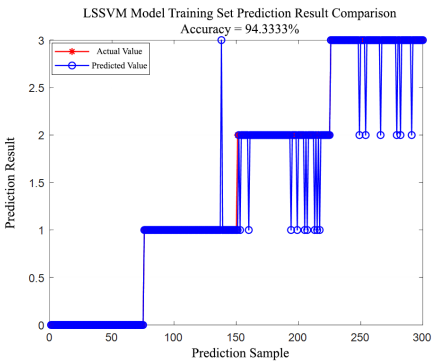
Fig. 7. Comparison of SVM recognition and classification results.



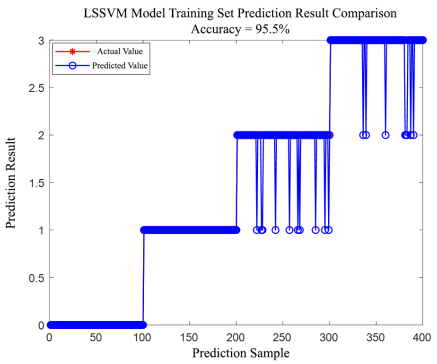
(a) 100 data



(b) 200 data

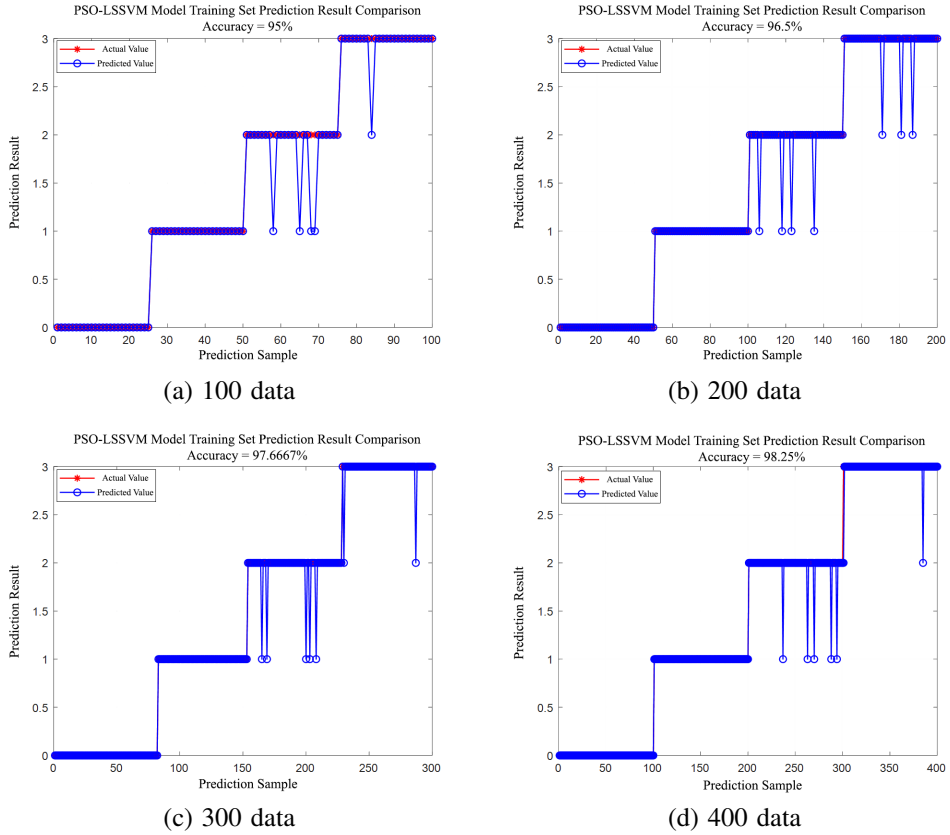


(c) 300 data



(d) 400 data

Fig. 8. Comparison of LSSVM recognition and classification results.



(a) 100 data (b) 200 data

(c) 300 data (d) 400 data

Fig. 9. Comparison of LSSVM recognition and classification results.

iteratively searches for the optimal hyperparameters of the LSSVM model. As shown in the figure, the fitness value drops markedly in the initial iterations. However, between the 7 and 28th iterations, the algorithm stagnates in a local optimum, with the fitness value remaining flat twice. By the 29th iteration, it escapes this local trap and converges to a global or near-global optimal solution, after which the fitness stabilizes without further reduction. Throughout this process, PSO dynamically adjusts LSSVM's regularization parameter C and radial basis function (RBF) kernel width parameter, ultimately identifying optimal or near-optimal combinations in the parameter space. This refinement directly enhances the model's performance by fine-tuning its decision boundaries to better capture the complex feature relationships underlying flow pattern classification.

To evaluate the performance of the PSO-LSSVM model in intelligent flash evaporation flow pattern recognition based on differential pressure, training set-test set comparison plots and confusion matrix analysis, as shown in Fig. 11, reveal strong validation results. The model achieves 98.5% accuracy on the training set and 97% on the test set, demonstrating robust generalization capability with no signs of overfitting. This provides a reliable machine learning solution for real-time monitoring and production optimization in flash evaporation processes across chemical and energy industries. By enabling precise flow pattern recognition, it assists engineers in adjusting operational parameters promptly to safeguard equipment efficiency and safety, while also offering novel technical insights for optimizing industrial flow pattern recognition models.

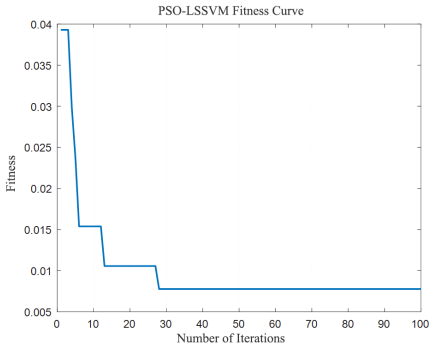


Fig. 10. Fitness Curve of PSO-LSSVM Model for Flow Pattern Recognition and Classification.

Analysis of the test set confusion matrix, paired with flow pattern encoding, shows distinct recognition performance across categories. Patterns 0 and 1 are classified with 100% accuracy, reflecting their stable differential pressure signal features and high distinguishability. The model reliably identifies these "mild flow patterns", providing accurate status assessments for stable production phases. For patterns 2 (90.5% accuracy) and 3 (95.8% accuracy), while minor misclassifications occur, overall performance remains strong. These more vigorous flow patterns, which characterized by complex

phase interfaces and variable flow structures, exhibit higher feature similarity in differential pressure signals, occasionally confusing the model. Nevertheless, their high recognition rates still effectively warn of "vigorous flow pattern" emergence, allowing timely engineer intervention to prevent equipment wear and productivity losses. This underscores the model's practical utility in differentiating flow patterns of varying intensity.

Analysis of the test set confusion matrix, paired with flow pattern encoding, shows distinct recognition performance across categories. Patterns 0 and 1 are classified with 100% accuracy, reflecting their stable differential pressure signal features and high distinguishability. The model reliably identifies these "mild flow patterns," providing accurate status assessments for stable production phases. For patterns 2 (90.5% accuracy) and 3 (95.8% accuracy), while minor misclassifications occur, overall performance remains strong. These more vigorous flow patterns, which characterized by complex phase interfaces and variable flow structures, exhibit higher feature similarity in differential pressure signals, occasionally confusing the model. Nevertheless, their high recognition rates still effectively warn of "vigorous flow pattern" emergence, allowing timely engineer intervention to prevent equipment wear and productivity losses. This underscores the model's practical utility in differentiating flow patterns of varying intensity.

4. Concluding discussion and remarks

This study advances flash evaporation flow pattern recognition in geothermal production wells by integrating experimental signal processing with targeted machine learning. We first established CEEMD as a reliable tool for extracting IMF energy spectra from differential pressure signals, which is critical for resolving the dynamics of phase change and two-phase flow. Subsequent investigations into sample design revealed that a dataset combining inlet temperature, flow velocity, and IMF energy features forms an optimal basis for model training, while PSO-LSSVM outperformed SVM and LSSVM by mitigating overfitting and enhancing generalization. Ultimately, the model's ability to distinguish flow patterns using multi-parameter features provides a robust technical foundation for locating flash evaporation initiation and assessing its severity in geothermal operations, addressing a key challenge in optimizing well performance and safety.

Future work should prioritize testing the PSO-LSSVM framework in actual downhole geothermal wells. Real-world conditions, including variable lithology, fluctuating production rates, and wellbore scaling, will challenge model robustness and guide refinements to bridge lab findings with industrial deployment. This step is critical to ensuring the method's utility in real production environments. Expanding beyond differential pressure signals could capture more nuanced features of ambiguous flow patterns. Multimodal fusion would likely improve recognition accuracy for transient or overlapping regimes, addressing current limitations and extending the model's applicability to complex flow scenarios.

Author information

The email addresses of the remaining author of this paper is as follows:

suns@tongji.edu.cn (S. Sun).

Acknowledgements

We would like to express appreciation to the following financial support: Shandong Excellent Young Scientist (Overseas) Program (No. 2024HWYQ-050) and Qingdao Postdoc Research Program (No. QDBSH20240201031).

Conflict of interest

The authors declare no competing interest.

References

Caraghiaur, D., Anglart, H. Drop deposition in annular two-phase flow calculated with Lagrangian Particle Tracking. Nuclear Engineering and Design, 2013, 265: 856-866.

Camargo, T. F. B., Paladino, E. E. Characterization of transient differential pressure signal features and flow pattern identification in horizontal two-phase flow through a constriction with machine learning models. Flow Measurement and Instrumentation, 2025, 106: 102985.

Dong, Y., Xue, L., Song, X., et al. Multi-physics coupling mechanisms and key development factors in enhanced geothermal systems for hot dry rock. Energy. 2025, 337: 138533.

Elgaddafi, R., Ahmed, R., Kiran, R., et al. Modeling and experimental studies on high-velocity slug and churn gas-liquid upward flows in vertical pipes. Geoenergy Science and Engineering, 2025, 257: 214210.

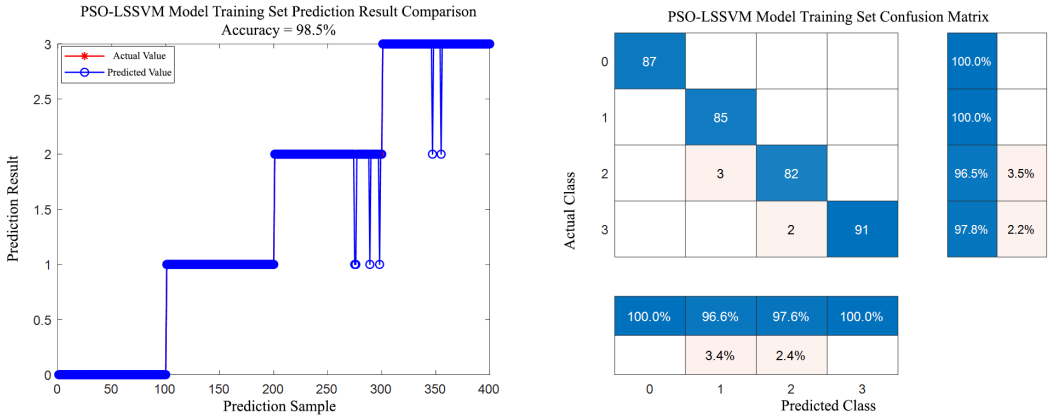
Gorjaei, R. G., Songolzadeh, R., Torkaman, M., et al. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. Journal of Natural Gas Science and Engineering, 2015, 24: 228-237.

Huang, N., Shen, Z., Long, S., et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.

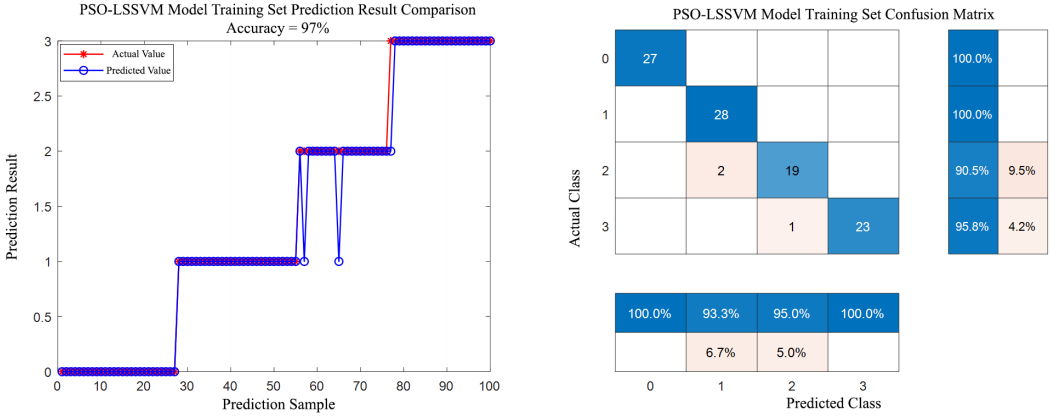
Liu, Y., Zhao, Q., Chong, D. Review of flow regime transition criteria for adiabatic co-current upward gas-liquid flow in vertical multi-scale channels. Measurement: Energy, 2025, 8: 100066.

Liu, W., Zhang, S., Chen, Z., et al. A fully coupled thermo-hydro-mechanical model for investigating heat extraction mechanisms in enhanced geothermal systems. Applied Thermal Engineering, 2025, 280: 128459.

Liu, W., Xu, Q., Guo, L. Pressure fluctuation characteristics and wavelet analysis of gas-liquid two-phase flow in gathering riser. Journal of Engineering Thermophysics, 2019, 40: 1820-1825. (in Chinese)



(a) Training Data



(b) Testing Data

Fig. 11. Recognition accuracy of PSO-LSSVM.

Marcel, C. P., Rohde, M., Van Der Hagen, T. H. Experimental investigations on flashing-induced instabilities in one and two-parallel channels: A comparative study. Experimental Thermal and Fluid Science, 2010, 34(7): 879-892.

Ma, T., Wang, T., Wang, L., et al. A hybrid deep learning model towards flow pattern identification of gas-liquid two-phase flows in horizontal pipe. Energy, 2025, 320: 135141.

Singh, M., Mahmoodpour, S., Singh, M., et al. Impact of thermo-hydro-mechanical process coupling for CO_2-water multiphase flow during geothermal energy extraction. Energy, 2025, 334: 137831

Singh, S., Chakraborty, P. R., Kothadia, H. B. Flash evaporation in a superheated liquid pool using water as medium. International Journal of Thermofluids, 2024, 24: 100939.

Tong, W., Zhu, G. Flow pattern recognition method of gas-liquid two-phase flow based on multilayer perceptron. Thermal Power Engineering, 2020, 35: 116-122. (in Chinese)

Yan, Y., Li, X. Ito, K. Numerical investigation of indoor particulate contaminant transport using the Eulerian-Eulerian and Eulerian-Lagrangian two-phase flow models. Experimental and Computational Multiphase Flow, 2020, 2: 31-40.

Wang, H. Prediction of gas-liquid two-phase holdup based on SOA-optimized LSSVM. Petroleum Engineering Construction, 2023, 49: 25-30. (in Chinese)

Wang, L., Li, H., Bu, X. Thermo-economic investigation of binary flashing cycle for enhanced geothermal system. Geothermics, 2021, 89: 101951.

Wu, Z., Huang, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in adaptive data analysis, 2009, 1(1): 1-41.

Yuan, S., Zhang, L., Li, T., et al. Study of Gas-Liquid Two-Phase Flow Characteristics at the Pore Scale Based on the VOF Model. Energies, 2025, 18(2): 316.

Zaborowska, I., Grzybowski, H., Mosdorf, R. Boiling flow pattern identification using a self-organizing map. Applied Sciences, 2020, 10(8): 2792.