Journal of Computational and Applied Mathematics 491 (2027) 118047

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A self-adaptive two-step inertial–viscosity algorithm for monotone inclusion problems with applications to medical image restoration and multi-class classification

Suthep Suantai [a,b], Preeyanuch Chuasuk [c], Watcharaporn Cholamjiak [d], Bing Tan [e], Pronpat Peeyada [f,g,*]

[a] *Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand*
[b] *Excellence Center in Optimization and Computational Intelligence for Big Data Prediction, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand*
[c] *Department of Mathematics, Faculty of Science, Burapha University, Chonburi, 20131, Thailand*
[d] *Advanced Numerical Optimization Research Group, Department of Mathematics, School of Science, University of Phayao, Phayao, 56000, Thailand*
[e] *School of Mathematics and Statistics, Southwest University, Chongqing, China*
[f] *School of Science, Walailak University, Nakhon Si Thammarat, 80160, Thailand*
[g] *Research Center for Theoretical Simulation and Applied Research in Bioscience and Sensing, Walailak University, Nakhon Si Thammarat, 80160, Thailand*

A R T I C L E I N F O

2000 MSC:
46N10
47H05
49M37
90C90

Keywords:
Inclusion problem
Fixed point problem
Image recovery
Data classification
Osteoporosis

A B S T R A C T

In this paper, we introduce a modification of Tseng's method with different step-size rules for solving an inclusion problem. The proposed algorithm integrates a two-step inertial mechanism together with a viscosity technique to enhance stability and convergence efficiency. Under suitable assumptions, including Lipschitz continuity and monotonicity conditions, we establish a strong convergence theorem for the generated iterative sequence. To support the theoretical findings, the proposed method is validated through numerical experiments on two different types of real-world datasets. First, in the image recovery task using medical X-ray images, the proposed method achieves superior structural preservation, visual clarity, and robustness in restoring fine anatomical details compared with related algorithms. Second, in the data classification task, the algorithm is embedded within an Extreme Learning Machine (ELM) framework and evaluated on both image-based and tabular datasets. The image-based dataset consists of multi-class knee osteoporosis X-ray images. Deep features extracted from a convolutional neural network are used as input to the ELM classifier. In addition, a structured tabular dataset is used to evaluate the algorithm's performance in numerical classification, with results indicating that the proposed optimization strategy enhances classification accuracy, improves precision–recall balance, accelerates convergence speed, and achieves better computational efficiency compared with existing methods.

* Corresponding author.
E-mail addresses: suthep.s@cmu.ac.th (S. Suantai), Preeyanuch.ch@buu.ac.th (P. Chuasuk), watcharaporn.ch@up.ac.th (W. Cholamjiak), bingtan@swu.edu.cn (B. Tan), pronpat.pe@wu.ac.th (P. Peeyada).

https://doi.org/10.1016/j.cam.2026.118047
Received 18 February 2026; Received in revised form 31 July 2026
Available online 9 August 2026

1. Introduction

Let $\mathbb{H}$ be a real Hilbert space equipped with an inner product $\langle .,. \rangle$ and the associated norm $\|.\|$. The inclusion problem can be formulated as:

$$\text{find } \hat{t} \in \mathbb{H} \text{ such that } 0 \in (X+Y)\hat{t}, \tag{1.1}$$

where $X : \mathbb{H} \to \mathbb{H}$ and $Y : \mathbb{H} \to 2^{\mathbb{H}}$ are single-valued and multi-valued mappings, respectively. The set of solution to problem (1.1) is denoted by $(X+Y)^{-1}(0)$. The inclusion problem (1.1) serves as a fundamental framework in optimization theory. It encompasses a broad class of mathematical problems, such as equilibrium problems, variational inequalities, convex minimization problems and split feasibility problems, see e.g. [1–6]. Furthermore, such models have been extensively utilized in various scientific and engineering fields, including signal and image processing, machine learning, linear inverse problem, transportation problem and engineering (see [7–11]).

Since the inclusion problem (1.1) is at the core of many mathematical problems [8,12–14], many researchers have proposed and developed iterative methods for solving the inclusion problem (1.1), but one of the most notable is the forward-backward splitting method introduced by Lions and Mercier [15], which is defined as follows:

$$t_{r+1} = (I+\vartheta Y)^{-1}(I-\vartheta X)t_r, \quad \forall r \geq 1, \tag{1.2}$$

where $\vartheta > 0$. In this setting (1.2), $(I-\vartheta X)$ is called the forward operator and $(I+\vartheta Y)^{-1}$ is called the backward operator. Under the assumption that the single-valued mapping X satisfies the co-coercive property and Y is maximally monotone, they proved that the sequences generated by the proposed algorithm converge weakly to a solution of the inclusion problem (1.1).

Lorenz and Pock [16], in 2015, studied problem (1.1) and introduced the inertial forward-backward algorithm. Their algorithm starts by choosing an arbitrary starting point $t_0, t_1 \in \mathbb{H}$ and generates $\{t_r\}$ be generated by

$$\begin{aligned} w_r &= t_r + \alpha_r(t_r - t_{r-1}), \\ t_{r+1} &= (I+\vartheta_r Y)^{-1}(I-\vartheta_r X)w_r, \quad \forall r \geq 1, \end{aligned} \tag{1.3}$$

where $\vartheta_1 > 0$, $\alpha_r \in [0,1)$ is an extrapolation factor, and the operators X and Y satisfy the assumptions required in [16]. The authors proved weak convergence result generated by Algorithm (1.3) to a solution point of problem (1.1) under certain conditions.

Gibali and Thong [17] studied the inclusion problem (1.1). They proposed a viscosity Tseng-type method, which combines the Tseng's method and viscosity-ideas. Their method generates a sequence $\{t_r\}$ as follows: Given an arbitrary starting point $t_0 \in \mathbb{H}$ and $\epsilon \in (0,1)$, compute

$$\begin{aligned} z_r &= (I+\vartheta_r Y)^{-1}(I-\vartheta_r X)t_r, \\ y_r &= z_r - \vartheta_r(Xz_r - Xt_r), \\ t_{r+1} &= \delta_r f(t_r) + (1-\delta_r)y_r, \end{aligned} \tag{1.4}$$

and $\{\vartheta_r\}$ is update by (1.5)

$$\vartheta_{r+1} = \begin{cases} \min\left\{ \dfrac{\epsilon\|t_r - z_r\|}{\|Xt_r - Xz_r\|}, \vartheta_r \right\}, & \text{if } Xt_r - Xz_r \neq 0; \\ \vartheta_r, & \text{otherwise,} \end{cases} \tag{1.5}$$

where $\vartheta_0 > 0$, f is a contraction with constant $\eta \in [0,1)$ and $\delta_r \in (0,1)$ satisfies $\lim_{r\to\infty} \delta_r = 0$ and $\sum_{r=1}^{\infty} \delta_r = \infty$. Under some mild assumptions, the sequences generated by Algorithm (1.4) converge strongly to an element $q \in \Omega$, where $q = P_{\Omega} \circ f(q)$.

Recently, Kaewyong and Sitthithakerngkiet [18] developed a modified iterative method that combines inertial and viscosity techniques based on Tseng's method, aimed at solving the inclusion problem (1.1). Let t_1 and $t_2 \in \mathbb{H}$ be arbitrary. Given $\vartheta_1 > 0$, $\epsilon \in (0,1)$, $\alpha > 0$ and $\{\delta_r\} \in (0,1)$ such that $\lim_{r\to\infty} \delta_r = 0$ and $\sum_{r=1}^{\infty} \delta_r = \infty$. Their algorithm generates $\{t_r\}$ by:

$$\begin{aligned} w_r &= t_r + \alpha_r(t_r - t_{r-1}), \\ z_r &= (I+\vartheta_r Y)^{-1}(I-\vartheta_r X)w_r, \\ y_r &= z_r - \vartheta_r(Xz_r - Xw_r), \\ t_{r+1} &= \delta_r f(t_r) + (1-\delta_r)y_r, \end{aligned} \tag{1.6}$$

where $\{\alpha_r\}$ and $\{\vartheta_r\}$ are update by (1.7) and (1.8), respectively

$$\alpha_r = \begin{cases} \min\left\{ \frac{\varepsilon_r}{\|t_r - t_{r-1}\|}, \alpha \right\}, & \text{if } t_{r-1} \neq t_r; \\ \alpha, & \text{otherwise,} \end{cases} \tag{1.7}$$

and

$$\vartheta_{r+1} = \begin{cases} \min\left\{ \dfrac{\epsilon\|t_r - z_r\|}{\|Xt_r - Xz_r\|}, \vartheta_r \right\}, & \text{if } Xt_r - Xz_r \neq 0; \\ \vartheta_r, & \text{otherwise,} \end{cases} \tag{1.8}$$

where f is an η-contraction $\eta \in [0,1)$, $\{\varepsilon_r\}$ is a positive sequence that satisfies $\lim_{r\to\infty} \frac{\varepsilon_r}{\delta_r} = 0$ and some suitable conditions. They proved a strong convergence result for Algorithm (1.6), which requires rather restrictive assumptions on X and Y, such that the single-valued mapping X is monotone and Lipschitz continuous and Y is maximal monotone.

Inspired by the aforementioned studies, we propose a new modified iterative method that uses a two-step inertial and a viscosity technique to solve inclusion problem in real Hilbert spaces. The aim of this new method is to develop a self-adaptive scheme that achieves strong convergence properties quickly under certain conditions, involving more general types of operators. Furthermore, we present some numerical examples to illustrate the computational effectiveness of our method in comparison to some existing ones in the literature. We also apply the proposed algorithm to solve image recovery and data classification problems.

The remainder of this article is structured as follows: In Section 2 presents the key definitions and lemmas used in the analysis of our algorithm. In Section 3 presents the algorithm is given along with essential assumptions and a convergence analysis of the main theorem. In Section 4 presents numerical examples demonstrating the effectiveness of the proposed method. Finally, we conduct experiments using the proposed algorithm to solve image recovery and data classification problems.

2. Preliminaries

In this section, we present some definition and lemmas, to prove our main theorem. Let $\{t_r\}$ be a sequence in $\mathbb{H}$. The strong convergence and weak convergence of $\{t_r\}$ to t are written by $t_r \to t$ and $t_r \rightharpoonup t$, respectively.

Definition 2.1. Let $X : \mathbb{H} \to \mathbb{H}$ be a mapping:

(i) The mapping X is called L-Lipschitz continuous with $L > 0$ if

$$\|Xt - Xs\| \le L\|t - s\|, \quad \forall t, s \in \mathbb{H}.$$

If the condition holds for $L = 1$ then X is called nonexpansive.

(ii) The mapping X is called firmly nonexpansive if

$$\|Xt - Xs\|^2 \le \langle Xt - Xs, t - s\rangle, \quad \forall t, s \in \mathbb{H},$$

or equivalently

$$\|Xt - Xs\|^2 \le \|t - s\|^2 - \|(I - X)t - (I - X)s\|^2, \quad \forall t, s \in \mathbb{H}.$$

(iii) The mapping X is called monotone if

$$\langle Xt - Xs, t - s\rangle \ge 0, \quad \forall t, s \in \mathbb{H}.$$

(iv) The fixed point set of X, denoted by $Fix(X)$ is defined as follows:

$$Fix(X) := \{t \in \mathbb{H} \mid Xt = t\}.$$

Definition 2.2. Let $Y : \mathbb{H} \to 2^{\mathbb{H}}$ be a multi-valued mapping. Then Y is said to be

(i) monotone if for all $(t,u), (s,v) \in graph(Y)$ (the graph of mapping Y),

$$\langle u - v, t - s\rangle \ge 0,$$

(ii) maximal monotone if for every $(t,u) \in \mathbb{H} \times \mathbb{H}$,

$$(t,u) \in graph(Y) \iff \langle u - v, t - s\rangle \ge 0, \ \forall (s,v) \in graph(Y).$$

Definition 2.3. Let $Y : \mathbb{H} \to 2^{\mathbb{H}}$ be a multi-valued maximal monotone mapping. The resolvent of Y, denoted by $J_\vartheta^Y : \mathbb{H} \to \mathbb{H}$, is defined by

$$J_\vartheta^Y(t) := (I + \vartheta Y)^{-1}(t), \quad \forall t \in \mathbb{H},$$

where $\vartheta > 0$ and I represents the identity mapping on $\mathbb{H}$.

The following lemmas are crucial to the convergence analysis of main results.

Lemma 2.4. *([19]) Let $Y : \mathbb{H} \to 2^{\mathbb{H}}$ be a set-valued maximal monotone mapping and $\vartheta > 0$. Then the domain of the resolvent of Y is the whole space, that is $D(J_\vartheta^Y) = \mathbb{H}$ and J_ϑ^Y is a single-valued and firmly nonexpansive mapping.*

Lemma 2.5. *([20]) Let $X : \mathbb{H} \to \mathbb{H}$ be a Lipschitz continuous and monotone mapping and $Y : \mathbb{H} \to 2^{\mathbb{H}}$ be a maximal monotone mapping. Then, the mapping $(X + Y)$ is a maximal monotone mapping.*

Lemma 2.6. *([21]) Let $\{\omega_r\}$ be a sequence of nonnegative real numbers and there exists $r_0 \in \mathbb{N}$ such that:*

$$\omega_{r+1} \leq (1-\varrho_r)\omega_r + \varrho_r \nu_r + \kappa_r, \quad \forall r \neq r_0,$$

where $\{\varrho_r\} \subset (0,1)$ and $\{\nu_r\}$, $\{\kappa_r\}$ satisfy the following conditions:

(i) $\sum_{r=0}^{\infty} \varrho_r = \infty$;
(ii) $\limsup_{r\to\infty} \nu_r \leq 0$;
(iii) $\kappa_r \geq 0 \ \ \forall r, \sum_{r=1}^{\infty} \kappa_r < \infty$.

Then, $\lim_{r\to\infty} \omega_r = 0$.

Lemma 2.7. *([22]) Let $\{\Pi_r\}$ be sequence of real numbers that does not decrease at infinity in the sense that, $\{\Pi_{r_j}\} \subset \{\Pi_r\}$ which satisfies $\Pi_{r_j} < \Pi_{r_j+1}$, for all $j \geq 1$. Let $\{\xi(r)\}_{r\geq r_0}$ be a sequence of integers defined as follows:*

$$\xi(r) = \max\{k \leq r \ : \ \Pi_k \leq \Pi_{k+1}\}. \tag{2.1}$$

Then $\{\xi(r)\}_{r\geq r_0}$ is a nondecreasing sequence with $\lim_{r\to\infty} \xi(r) = \infty$. Moreover, for each $r \geq r_0$, we have $\Pi_{\xi(r)} \leq \Pi_{\xi(r)+1}$ and $\Pi_r \leq \Pi_{\xi(r)+1}$.

3. Main results

In this section, we present a modified two-step inertial viscosity algorithm with a self-adaptive step size that generates a strongly convergent sequence for solving the inclusion problem (1.1). The following assumptions are used to analyze our method.

Assumption 3.1. We assume that:

(i) The solution set Ψ of (1.1) is nonempty.
(ii) $X : \mathbb{H} \to \mathbb{H}$ is L-Lipschitz continuous and a monotone mapping.
(iii) $Y : \mathbb{H} \to 2^{\mathbb{H}}$ is a maximal monotone mapping.
(iv) $f : \mathbb{H} \to \mathbb{H}$ is η-contraction with constant $\eta \in [0,1)$.

Assumption 3.2. Let $\{\delta_r\} \subset (0,1)$, $\{\rho_r\} \subset [1,\infty)$, $\{\varepsilon_r\} \subset [1,\infty)$ and $\{\varpi_r\} \subset [1,\infty)$ satisfy the following conditions:

(i) $\lim_{r\to\infty} \delta_r = 0$ and $\sum_{r=1}^{\infty} \delta_r = \infty$.
(ii) $\lim_{r\to\infty} \rho_r = 1$.
(iii) $\lim_{r\to\infty} \frac{\varepsilon_r}{\delta_r} = 0$.
(iv) $\lim_{r\to\infty} \frac{\varpi_r}{\delta_r} = 0$.

Remark 3.3. From (3.1), (3.2) and Assumption 3.2, we observe that

(i) $\lim_{r\to\infty} \alpha_r\|t_r - t_{r-1}\| = 0$ and $\lim_{r\to\infty} \frac{\alpha_r}{\delta_r}\|t_r - t_{r-1}\| = 0$.
(ii) $\lim_{r\to\infty} \beta_r\|t_{r-1} - t_{r-2}\| = 0$ and $\lim_{r\to\infty} \frac{\beta_r}{\delta_r}\|t_{r-1} - t_{r-2}\| = 0$.

Lemma 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then the sequence $\{\vartheta_r\}$ generated by (3.3) is well defined and $\lim_{r\to\infty} \lambda_r = \epsilon$, where $\lambda_r = \frac{\epsilon\rho_r\vartheta_r}{\vartheta_{r+1}}$.*

Proof. Since X is Lipschitz continuous with $L > 0$ and $\rho_r \geq 1$, if $Xw_r - Xz_r \neq 0$, then

$$\frac{\epsilon\rho_r\|w_r - z_r\|}{\|Xw_r - Xz_r\|} \geq \frac{\epsilon\rho_r\|w_r - z_r\|}{L\|w_r - z_r\|} = \frac{\epsilon\rho_r}{L} \geq \frac{\epsilon}{L}.$$

By using the same technique as in the proof of [Lemma 3.1, [23]], we obtain

$$\lim_{r\to\infty} \vartheta_r = \vartheta \in \left[\min\left\{\vartheta_1, \frac{\epsilon}{L}\right\}, \vartheta_1 + \sigma\right],$$

where $\sigma = \sum_{r=1}^{\infty} \sigma_r$. Therefore,

$$\lim_{r\to\infty} \lambda_r = \lim_{r\to\infty} \frac{\epsilon\rho_r\vartheta_r}{\vartheta_{r+1}} = \epsilon.$$

□

Algorithm 1: A self adaptive two-step inertial viscosity algorithm.

Initialization: Given $\alpha > 0$, $\beta > 0$, $\vartheta_1 > 0$, $\epsilon \in (0,1)$. Let $t_{-1}, t_0, t_1 \in \mathbb{H}$ be arbitrary. Choose $\{\sigma_r\} \subset [0,\infty)$ such that $\sum_{r=1}^{\infty} \sigma_r < \infty$ and set $r := 1$.
Iterative steps: Calculate the next iterate t_{r+1} as follows:
Step 1. Compute

$$\begin{aligned} w_r &= t_r + \alpha_r(t_r - t_{r-1}) + \beta_r(t_{r-1} - t_{r-2}), \\ z_r &= J^Y_{\vartheta_r}(I - \vartheta_r X)w_r, \\ y_r &= z_r - \vartheta_r(Xz_r - Xw_r), \end{aligned}$$

and

$$t_{r+1} = \delta_r f(y_r) + (1-\delta_r)y_r,$$

where

$$\alpha_r = \begin{cases} \min\left\{\frac{\epsilon_r}{\|t_r - t_{r-1}\|}, \alpha\right\}, & \text{if } \ t_{r-1} \neq t_r; \\ \alpha, & \text{otherwise}, \end{cases} \tag{3.1}$$

and

$$\beta_r = \begin{cases} \min\left\{\frac{\varpi_r}{\|t_{r-1} - t_{r-2}\|}, \beta\right\}, & \text{if } \ t_{r-2} \neq t_{r-1}; \\ \beta, & \text{otherwise}. \end{cases} \tag{3.2}$$

Step 2. Update

$$\vartheta_{r+1} = \begin{cases} \min\left\{\dfrac{\epsilon\rho_r\|w_r - z_r\|}{\|Xw_r - Xz_r\|}, \vartheta_r + \sigma_r\right\}, & \text{if } \ Xw_r - Xz_r \neq 0; \\ \vartheta_r + \sigma_r, & \text{otherwise}. \end{cases} \tag{3.3}$$

Set $r := r+1$ and go to **Step 1**.

Lemma 3.5. *Assume that Assumptions 3.1 and 3.2 hold and let a sequence $\{t_r\}$ be generated by Algorithm 1. Then, we have*

$$\|y_r - q\|^2 \le \|w_r - q\|^2 - (1 - \epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2})\|w_r z_r\|^2,$$

$\forall q \in \Psi$ and

$$\|y_r - z_r\| \le \epsilon\rho_r\frac{\vartheta_r}{\vartheta_{r+1}}\|w_r - z_r\|.$$

Proof. First, using the definition of $\{\vartheta_r\}$, we establish the following result

$$\|Xw_r - Xz_r\| \le \frac{\epsilon\rho_r}{\vartheta_{r+1}}\|w_r - z_r\|, \ \forall r \in \mathbb{N}. \tag{3.4}$$

Indeed, if $Xw_r = Xz_r$ then (3.4) clearly holds. Otherwise, it follows from (3.3) that

$$\vartheta_{r+1} = \min\left\{\frac{\epsilon\rho_r\|w_r - z_r\|}{\|Xw_r - Xz_r\|}, \vartheta_r + \sigma_r\right\} \le \frac{\epsilon\rho_r\|w_r - z_r\|}{\|Xw_r - Xz_r\|}.$$

Consequently, we have

$$\|Xw_r - Xz_r\| \le \frac{\epsilon\rho_r}{\vartheta_{r+1}}\|w_r - z_r\|.$$

Therefore, inequality (3.4) hold when $Xw_r = Xz_r$ and $Xw_r \neq Xz_r$. Let $q \in \Psi$, we get

$$\begin{aligned} \|y_r - q\|^2 &= \|z_r - \vartheta_r(Xz_r - Xw_r) - q\|^2 \\ &= \|z_r - q\|^2 + \vartheta_r^2\|Xz_r - Xw_r\|^2 - 2\vartheta_r\langle z_r - q, Xz_r - Xw_r\rangle \\ &= \|z_r - w_r\|^2 + \|w_r - q\|^2 + 2\langle z_r - w_r, w_r - q\rangle + \vartheta_r^2\|Xz_r - Xw_r\|^2 - 2\vartheta_r\langle z_r - q, Xz_r - Xw_r\rangle \\ &= \|w_r - q\|^2 - \|z_r - w_r\|^2 + 2\langle z_r - w_r, z_r - q\rangle + \vartheta_r^2\|Xz_r Xw_r\|^2 - 2\vartheta_r\langle z_r - q, Xz_r - Xw_r\rangle \\ &= \|w_r - q\|^2 - \|z_r - w_r\|^2 + \vartheta_r^2\|Xz_r - Xw_r\|^2 - 2\langle w_r - z_r - \vartheta_r(Xw_r - Xz_r), z_r - q\rangle. \end{aligned} \tag{3.5}$$

By using (3.4) and (3.5), we have

$$\begin{aligned}\|y_r - q\|^2 &\leq \|w_r - q\|^2 - \|z_r - w_r\|^2 + \vartheta_r^2\Big(\frac{\epsilon^2\rho_r^2}{\vartheta_{r+1}^2}\|w_r - z_r\|^2\Big) - 2\langle w_r - z_r - \vartheta_r(Xw_r - Xz_r), z_r - q\rangle \\ &= \|w_r - q\|^2 - \Big(1 - \epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\Big)\|w_r - z_r\|^2 - 2\langle w_r - z_r - \vartheta_r(Xw_r - Xz_r), z_r - q\rangle. \end{aligned} \tag{3.6}$$

Since $z_r = (I + \vartheta_r Y)^{-1}(I - \vartheta_r X)w_r$ and Y is maximal monotone, we obtain

$$\frac{w_r - \vartheta_r Xw_r - z_r}{\vartheta_r} \in Yz_r,$$

which implies that

$$\frac{w_r - \vartheta_r Xw_r - z_r}{\vartheta_r} + Xz_r \in (X+Y)z_r.$$

Since $X : \mathbb{H} \to \mathbb{H}$ is monotone and $Y : \mathbb{H} \to 2^{\mathbb{H}}$ is a maximal monotone, thus by Lemma 2.5, we have that $(X+Y)$ is maximal monotone. So, since $q \in \Psi$, then $0 \in (X+Y)q$, thus

$$\langle w_r - z_r - \vartheta_r(Xw_r - Xz_r), z_r - q\rangle \geq 0. \tag{3.7}$$

By using (3.7) in (3.6), we obtain

$$\|y_r - q\|^2 \leq \|w_r - q\|^2 - \Big(1 - \epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\Big)\|w_r - z_r\|^2.$$

According to the definition of y_r and (3.4), we have

$$\|y_r - z_r\| \leq \epsilon\rho_r\frac{\vartheta_r}{\vartheta_{r+1}}\|w_r - z_r\|. \tag{3.8}$$

Thus, the proof of Lemma 3.5 is complete. □

Theorem 3.6. *Assume that Assumptions 3.1 and 3.2 hold. Let a sequence $\{t_r\}$ be generated by Algorithm 1. Then, $\{t_r\}$ converges strongly to $q \in \Psi$, where $q = P_\Psi f(q)$.*

Proof. Firstly, we prove that $\{t_r\}$ is a bounded sequence. According to Lemma 3.5, we get that $\lim_{r\to\infty}\Big(1 - \epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\Big) = 1 - \epsilon^2 > 0$. Therefore, there is a constant $r_0 \in \mathbb{N}$ that satisfies $1 - \epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2} > 0$, $\forall r \geq r_0$. Thus, from Lemma 3.5, we obtain

$$\|y_r - q\| \leq \|w_r - q\|, \ \forall r \geq r_0. \tag{3.9}$$

By the definition of w_r, we have

$$\begin{aligned}\|w_r - q\| &= \|t_r + \alpha_r(t_r - t_{r-1}) + \beta_r(t_{r-1} - t_{r-2}) - q\| \\ &\leq \|t_r - q\| + \alpha_r\|t_r - t_{r-1}\| + \beta_r\|t_{r-1} - t_{r-2}\| \\ &= \|t_r - q\| + \delta_r \cdot \frac{\alpha_r}{\delta_r}\|t_r - t_{r-1}\| + \delta_r \cdot \frac{\beta_r}{\delta_r}\|t_{r-1} - t_{r-2}\|. \end{aligned} \tag{3.10}$$

Since, by Remark 3.3, $\lim_{r\to\infty}\frac{\alpha_r}{\delta_r}\|t_r - t_{r-1}\| = 0$ and $\lim_{r\to\infty}\frac{\beta_r}{\delta_r}\|t_{r-1} - t_{r-2}\| = 0$. Thus, there exists a constant $M_1 > 0$ such that

$$\frac{\alpha_r}{\delta_r}\|t_r - t_{r-1}\| \leq M_1, \quad \forall r \geq 1, \tag{3.11}$$

and $M_2 > 0$ such that

$$\frac{\beta_r}{\delta_r}\|t_{r-1} - t_{r-2}\| \leq M_2, \quad \forall r \geq 1. \tag{3.12}$$

Combining (3.9), (3.10), (3.11) and (3.12), we obtain

$$\|y_r - q\| \leq \|w_r - q\| \leq \|t_r - q\| + \delta_r(M_1 + M_2). \tag{3.13}$$

Using the definition of $\{t_{r+1}\}$, we have for all $r \geq r_0$

$$\begin{aligned}\|t_{r+1} - q\| &= \|\delta_r f(y_r) + (1-\delta_r)y_r - q\| \\ &\leq \delta_r\|f(y_r) - q\| + (1-\delta_r)\|y_r - q\| \\ &\leq \delta_r\|f(y_r) - f(q)\| + \delta_r\|f(q) - q\| + (1-\delta_r)\|y_r - q\| \\ &\leq \eta\delta_r\|y_r - q\| + \delta_r\|f(q) - q\| + (1-\delta_r)\|y_r - q\| \end{aligned}$$

$$
\begin{aligned}
&= (1-(1-\eta)\delta_r)\|y_r - q\| + \delta_r\|f(q)-q\| \\
&\le (1-(1-\eta)\delta_r)\|t_r - q\| + (1-\eta)\delta_r\left[\frac{M_1+M_2+\|f(q)-q\|}{1-\eta}\right] \\
&\le \max\left\{\|t_r-q\|, \frac{M_1+M_2+\|f(q)-q\|}{1-\eta}\right\} \\
&\vdots \\
&\le \max\left\{\|t_{r_0}-q\|, \frac{M_1+M_2+\|f(q)-q\|}{1-\eta}\right\}.
\end{aligned}
$$

Thus, $\{t_r\}$ is bounded. Consequently $\{w_r\}$, $\{y_r\}$ and $\{f(y_r)\}$ are also bounded.
Next, we observe that

$$
\begin{aligned}
\|w_r - q\|^2 &\le [\|t_r - q\| + \delta_r(M_1+M_2)]^2 \\
&= \|t_r-q\|^2 + 2\delta_r(M_1+M_2)\|t_r-q\| + \delta_r^2(M_1+M_2)^2 \\
&= \|t_r-q\|^2 + \delta_r\left[2(M_1+M_2)\|t_r-q\| + \delta_r(M_1+M_2)^2\right] \\
&\le \|t_r-q\|^2 + \delta_r M_3,
\end{aligned} \tag{3.14}
$$

for some $M_3 > 0$. Combining Lemma 3.5 and (3.14), we see that

$$
\begin{aligned}
\|t_{r+1}-q\|^2 &\le \delta_r\|f(y_r)-q\|^2 + (1-\delta_r)\|y_r-q\|^2 \\
&\le \|y_r-q\|^2 + \delta_r(\|f(y_r)-q\|^2 + 2\|y_r-q\|\|f(q)-q\|) \\
&\le \|y_r-q\|^2 + \delta_r M_4 \\
&\le \|w_r-q\|^2 - \left(1-\epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\right)\|w_r-z_r\|^2 + \delta_r M_4 \\
&\le \|t_r-q\|^2 + \delta_r M_3 - \left(1-\epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\right)\|w_r-z_r\|^2 + \delta_r M_4 \\
&= \|t_r-q\|^2 - \left(1-\epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\right)\|w_r-z_r\|^2 + \delta_r M_5,
\end{aligned}
$$

where $M_5 := M_3 + M_4$. Therefore, we obtain

$$
\left(1-\epsilon^2\rho_r^2\frac{\vartheta_r^2}{\vartheta_{r+1}^2}\right)\|w_r-z_r\|^2 \le \|t_r-q\|^2 - \|t_{r+1}-q\|^2 + \delta_r M_5. \tag{3.15}
$$

Using the definition of w_r, we can show that

$$
\begin{aligned}
\|w_r-q\|^2 &= \|t_r + \alpha_r(t_r-t_{r-1}) + \beta_r(t_{r-1}-t_{r-2}) - q\|^2 \\
&\le \|t_r-q\|^2 + 2\alpha_r\|t_r-q\|\|t_r-t_{r-1}\| + \alpha_r^2\|t_r-t_{r-1}\|^2 \\
&\quad + 2\beta_r\|t_r+\alpha_r(t_r-t_{r-1})-q\|\|t_{r-1}-t_{r-2}\| + \beta_r^2\|t_{r-1}-t_{r-2}\|^2 \\
&= \|t_r-q\|^2 + \alpha_r\|t_r-t_{r-1}\|(2\|t_r-q\| + \alpha_r\|t_r-t_{r-1}\|) \\
&\quad + \beta_r\|t_{r-1}-t_{r-2}\|(2\|t_r+\alpha_r(t_r-t_{r-1})-q\| + \beta_r\|t_{r-1}+t_{r-2}\|) \\
&\le \|t_r-q\|^2 + \alpha_r\|t_r-t_{r-1}\|M + \beta_r\|t_{r-1}-t_{r-2}\|M,
\end{aligned} \tag{3.16}
$$

where $M := \sup_{r\in\mathbb{N}}\left\{2\|t_r-q\| + \alpha_r\|t_r-t_{r-1}\|, 2\|t_r+\alpha_r(t_r-t_{r-1})-q\| + \beta_r\|t_{r-1}+t_{r-2}\|\right\} < \infty$. Using (3.9) and (3.16), we have for all $r \ge r_0$

$$
\begin{aligned}
\|t_{r+1}-q\|^2 &= \|\delta_r f(y_r) + (1-\delta_r)y_r - q\|^2 \\
&\le \|\delta_r(f(y_r)-f(q)) + (1-\delta_r)(y_r-q)\|^2 + 2\delta_r\langle f(q)-q, \delta_r f(y_r) + (1-\delta_r)y_r - q\rangle \\
&\le \delta_r\|f(y_r)-f(q)\|^2 + (1-\delta_r)\|y_r-q\|^2 + 2\delta_r\langle f(q)-q, t_{r+1}-q\rangle \\
&\le \eta^2\delta_r\|y_r-q\|^2 + (1-\delta_r)\|y_r-q\|^2 + 2\delta_r\langle f(q)-q, t_{r+1}-q\rangle \\
&= (1-(1-\eta^2)\delta_r)\|y_r-q\|^2 + 2\delta_r\langle f(q)-q, t_{r+1}-q\rangle \\
&\le (1-(1-\eta^2)\delta_r)\|w_r-q\|^2 + 2\delta_r\langle f(q)-q, t_{r+1}-q\rangle \\
&\le (1-(1-\eta^2)\delta_r)\|t_r-q\|^2 + (1-\eta^2)\delta_r\left[\frac{3M}{(1-\eta^2}\left(\frac{\alpha_r}{\delta_r}\|t_r-t_{r-1}\| + \frac{\beta_r}{\delta_r}\|t_{r-1}-t_{r-2}\|\right)\right. \\
&\quad \left. + \frac{2}{1-\eta^2}\langle f(q)-q, t_{r+1}-q\rangle\right].
\end{aligned} \tag{3.17}
$$

Next, we consider the following two cases.
Case 1. We can find $N \in \mathbb{N}$ satisfying $\Pi_{r+1} \leq \Pi_r$, for each $r \geq N$, where $\Pi_r = \|t_r - q\|^2$. Since each term Π_r is nonnegative, it is convergent. Due to the fact that $\lim_{r\to\infty} \delta_r = 0$ and $\lim_{r\to\infty} \left(1 - \epsilon^2 \rho_r^2 \frac{\vartheta_r^2}{\vartheta_{r+1}^2}\right) > 0$, we obtain from (3.15) that

$$\lim_{r\to\infty} \|w_r - z_r\| = 0. \tag{3.18}$$

Thus, from (3.8), we immediately obtain

$$\lim_{r\to\infty} \|y_r - z_r\| = 0.$$

If we consider

$$\|y_r - w_r\| \leq \|y_r - z_r\| + \|w_r - z_r\|,$$

then

$$\lim_{r\to\infty} \|y_r - w_r\| = 0.$$

Since $\{t_r\}$ is bounded, there exists a subsequence $\{t_{r_k}\}$ of $\{t_r\}$ such that $t_{r_k} \rightharpoonup q^* \in \mathbb{H}$. By setting $T_r = J^Y_{\vartheta_r}(I - \vartheta_r X)$, we have

$$\begin{aligned}\|(I - T_r)q^*\|^2 &= \langle (I - T_r)q^*, (I - T_r)q^*\rangle \\ &= \langle (I - T_r)q^*, q^* - w_{r_k}\rangle + \langle (I - T_r)q^*, w_{r_k} - T_r w_{r_k}\rangle + \langle (I - T_r)q^*, T_r w_{r_k} - T_r q^*\rangle\end{aligned}$$

Using the fact that $\|t_r - w_r\| \to 0$ and (3.18), we obtain

$$\lim_{r\to\infty} \|(I - T_r)q^*\| = 0.$$

Therefore, $q^* \in \Psi$. We can obtain

$$\begin{aligned}\limsup_{r\to\infty} \frac{2}{1-\eta^2}\langle f(q) - q, t_{r+1} - q\rangle &= \limsup_{k\to\infty} \frac{2}{1-\eta^2}\langle f(q) - q, t_{r_k} - q\rangle \\ &= \frac{2}{1-\eta^2}\langle f(q) - q, q^* - q\rangle \\ &\leq 0.\end{aligned} \tag{3.19}$$

Applying Lemma 2.6 to the inequality from (3.17), we can conclude that $\lim_{r\to\infty} \Pi_r = 0$.
Case 2. We can find $r_k \in \mathbb{N}$ satisfying $r_k \geq k$ and $\Pi_{r_k} < \Pi_{r_k+1}$ for all $k \in \mathbb{N}$. According to Lemma 2.7, the inequality $\Pi_{\xi(r)} \leq \Pi_{\xi(r)+1}$ is obtained, where $\xi : \mathbb{N} \to \mathbb{N}$ is defined by (2.1), and $r \geq r^*$ for some $r^* \in \mathbb{N}$. This implies, by (3.15), for each $r \geq r^*$, as follows

$$\left(1 - \epsilon^2 \rho_{\xi(r)}^2 \frac{\vartheta_{\xi(r)}^2}{\vartheta_{\xi(r)+1}^2}\right)\|w_{\xi(r)} - z_{\xi(r)}\|^2 \leq \Pi_{\xi(r)} - \Pi_{\xi(r)+1} + \delta_{\xi(r)} M_5.$$

Similar to Case 1, since $\delta_r \to 0$ as $r \to \infty$, we obtain

$$\lim_{r\to\infty} \|w_{\xi(r)} - z_{\xi(r)}\| = 0.$$

Furthermore, an argument similar to the one used in Case a shows that

$$\limsup_{r\to\infty} \langle f(q) - q, t_{\xi(r)+1} - q\rangle \leq 0.$$

Finally, from the inequality $\Pi_{\xi(r)} \leq \Pi_{\xi(r)+1}$ and by (3.17), for all $r \geq \max\{r_0, r^*\}$, we obtain

$$\begin{aligned}\Pi_{\xi(r)+1} &\leq (1 - (1-\eta^2)\delta_{\xi(r)})\Pi_{\xi(r)} + (1-\eta^2)\delta_{\xi(r)}\left[\frac{2}{1-\eta^2}\langle f(q) - q, t_{\xi(r)+1} - q\rangle\right. \\ &\quad \left. + \frac{3M}{(1-\eta^2)}\left(\frac{\alpha_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)} - t_{\xi(r)-1}\| + \frac{\beta_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)-1} - t_{\xi(r)-2}\|\right)\right] \\ &\leq (1 - (1-\eta^2)\delta_{\xi(r)})\Pi_{\xi(r)+1} + (1-\eta^2)\delta_{\xi(r)}\left[\frac{2}{1-\eta^2}\langle f(q) - q, t_{\xi(r)+1} - q\rangle\right. \\ &\quad \left. + \frac{3M}{(1-\eta^2)}\left(\frac{\alpha_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)} - t_{\xi(r)-1}\| + \frac{\beta_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)-1} - t_{\xi(r)-2}\|\right)\right].\end{aligned}$$

Some simple calculations yield

$$\Pi_{\xi(r)+1} \leq \frac{3M}{(1-\eta^2)}\left(\frac{\alpha_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)} - t_{\xi(r)-1}\| + \frac{\beta_{\xi(r)}}{\delta_{\xi(r)}}\|t_{\xi(r)-1} - t_{\xi(r)-2}\|\right) + \frac{2}{1-\eta^2}\langle f(q) - q, t_{\xi(r)+1} - q\rangle.$$

From this it follows that $\limsup_{r\to\infty} \Pi_{\xi(r)+1} \leq 0$. Thus, $\lim_{r\to\infty} \Pi_{\xi(r)+1} = 0$. In addition, by Lemma 2.7,

$$\lim_{r\to\infty} \Pi_r \leq \lim_{r\to\infty} \Pi_{\xi(r)+1} = 0.$$

Hence, we can conclude that t_r converges strongly to q. □

Table 1
Numerical comparison results of each algorithm.

		Algorithm 1.3	Algorithm 1.4	Algorithm 1.6	Algorithm 1
Case I	No. of Iter.	249	265	236	93
	CPU time (s)	0.0216	0.0254	0.0117	0.0089
Case II	No. of Iter.	362	374	203	81
	CPU time (s)	0.0156	0.0234	0.0154	0.0078
Case III	No. of Iter.	191	201	158	65
	CPU time (s)	0.0170	0.0204	0.0152	0.0071
Case IV	No. of Iter.	169	180	147	61
	CPU time (s)	0.0151	0.0186	0.0141	0.0066

4. Numerical example

In this section, we present some numerical experiment to illustrate the performance of Algorithm 1. We compare our proposed algorithm with the Algorithms 1.3–1.6. We used MATLAB 2025a on a personal computer with 8 GB of RAM to perform all numerical calculations.

Example 4.1. Consider the inclusion problem (1.1) in an infinite-dimensional space $L_2([0,2\pi])$, where $L_2([0,2\pi]) = \left\{t = t(x) \in \mathbb{R}, t \in [0,2\pi] : \int_0^{2\pi} |t(x)|^2 dx < \infty\right\}$ with inner product $\langle t, s\rangle = \int_0^{2\pi} t(x)s(x)dx$ and norm $\|t\|_{L_2} = \left(\int_0^{2\pi} |t(x)|^2 dx\right)^{\frac{1}{2}}, \forall t, s \in L_2([0,2\pi])$. We define the mapping $X : L_2([0,2\pi]) \to L_2([0,2\pi])$ by $(Xt)(x) = t(x), \ \forall t \in L_2([0,2\pi])$ and $Y : L_2([0,2\pi]) \to L_2([0,2\pi])$ by $(Yt)(x) = 2t(x), \ \forall t \in L_2([0,2\pi])$.

In this experiment, we chose the following control parameters: $\alpha = 0.45$, $\epsilon_r = \frac{1}{(1000r+1)^3}$, $\vartheta_0 = \vartheta_1 = 0.02$, $\delta_r = \frac{1}{1000r+1}$, $\epsilon = 0.75$ and $f(t) = \frac{t}{2}$. In Algorithm 1 we choose $\beta = 0.25$, $\varpi_r = \frac{1}{(1000r+1)^3}$, $\rho_r = 1 + \frac{1}{r+1}$ and $\sigma_r = \frac{1}{(5r+2)^2}$. For Algorithm 1.3, we set $\alpha_r = \frac{1}{r+1}$. We choose different initial values, divided into the following four cases:

Case I: $t_{-1} = x\exp(-x)$, $t_0 = 5x\exp(x)$ and $t_1 = x^4 + 5x^2 + 2x - 1$.
Case II: $t_{-1} = \frac{x^2-\exp(-2x)}{5}$, $t_0 = \frac{x^3\exp(2x)}{5}$ and $t_1 = 1 - 2x + x^3$.
Case III: $t_{-1} = \frac{\exp(-2x)-x^3+1}{2}$, $t_0 = \frac{2x^3+5x-1}{2}$ and $t_1 = \frac{-x^2\sin(4x)}{2}$.
Case IV: $t_{-1} = 5x^3 + 2x^2 + x - 1$, $t_0 = x^2 + 3x + 1$ and $t_1 = 10x^2\exp(-x)$.

The stopping criterion used in the experiment is $\|t_{r+1} - t_r\| < 10^{-4}$, or when the number of iterations reaches the prescribed maximum of 1000. The numerical results of the example is given in Table 1 and Fig. 1.

Remark 4.2. From the results obtained in Table 1 and Fig. 1, it can be seen that in all four cases, Algorithm 1, when used as two step inertia, has better performance in terms of the number of iterations and the least time compared to the other three algorithms.

5. Application

This section presents a series of numerical experiments to assess the effectiveness of Algorithm 1 in solving the image restoration and data classification problems, in comparison with the three algorithms described earlier.

5.1. Image recovery problem

In this subsection, we will apply it to image recovery. The image recovery problem can be formulated as a system of linear equations as follows:

$$b = \mathcal{A}t + \varsigma, \tag{5.1}$$

where $t \in \mathbb{R}^{n\times 1}$ is an original image, $b \in \mathbb{R}^{m\times 1}$ is an observed image, $\varsigma \in \mathbb{R}^{m\times 1}$ is additive noise and $\mathcal{A} \in \mathbb{R}^{m\times n}$ is a blurring matrix. It has been known that solving (5.1) can be seen as the following convex minimization problem:

$$\min_{t\in\mathbb{R}^N} \frac{1}{2}\|b - \mathcal{A}t\|_2^2 + \tau\|t\|_1, \tag{5.2}$$

where $\tau > 0$ is a regularization parameter. The problem (5.2) can be transformed into the inclusion problem (1.1) by setting $X(t) \equiv \nabla\left(\frac{1}{2}\|\mathcal{A}t - b\|^2\right)$ and $Y(t) \equiv \partial(\tau\|t\|_1)$.

We use two well-known metrics to measure the quality of the recovered images, namely the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM), where PSNR is defined as follows:

$$\text{PSNR} := 20\log_{10}\left(\frac{255^2}{\|t_r - t\|_2^2}\right),$$

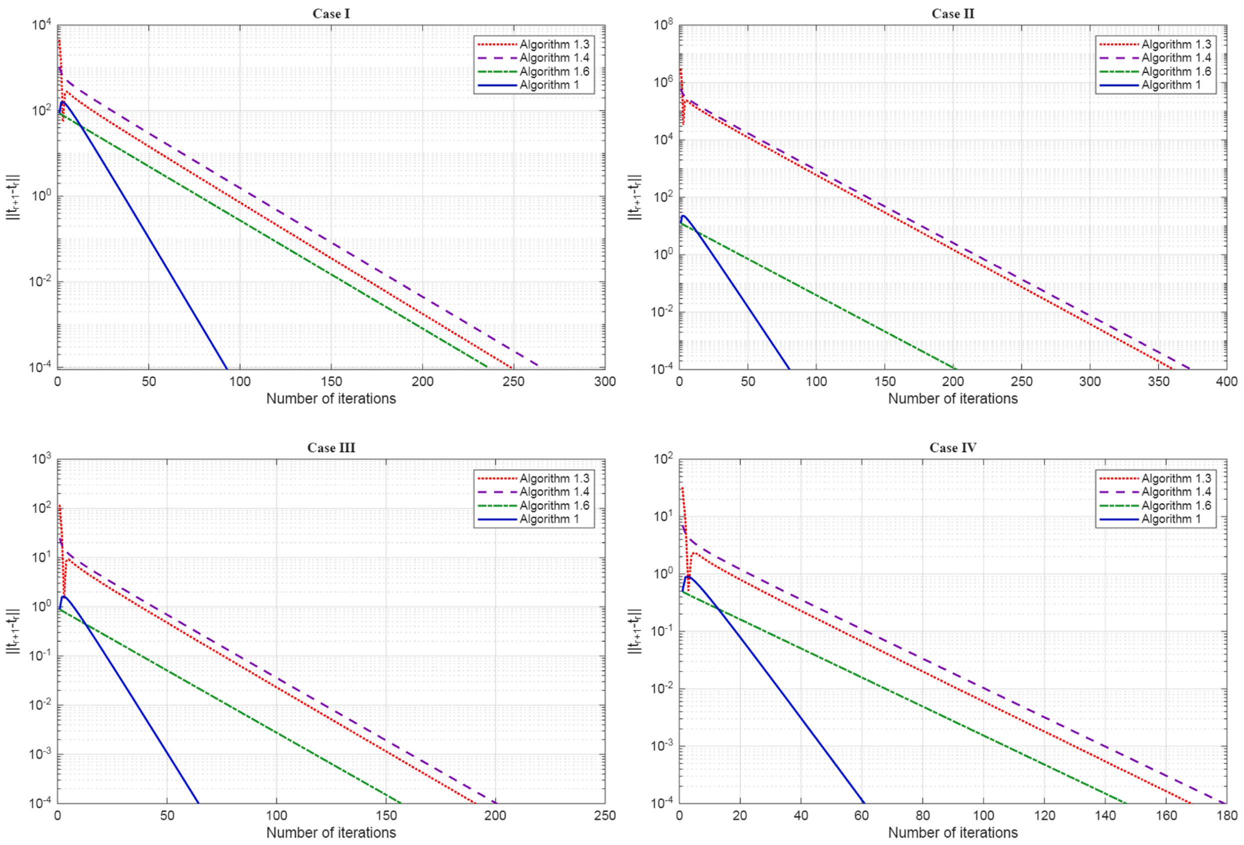

Fig. 1. Error plotting of $\|t_{r+1} - t_r\|$ for the different cases I – IV.

where t is an original image and t_r is a restored image. The SSIM is defined by

$$\mathrm{SSIM} := \frac{(2\psi_x\psi_y + C_1)(2\sigma_{xy} + C_2)}{(\psi_x^2 + \psi_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where ψ_x and ψ_y are the mean values of the original and restored image, respectively. The variances are symbolized as σ_x^2 and σ_y^2. The covariance between the images as σ_{xy}. We also denote $C_1 = (0.01S)^2$, $C_2 = (0.03S)^2$ and S is the dynamic range of pixel values in the image. The parameters of each algorithm are set as follows.

- For Algorithm 1.3, we set $\vartheta_1 = 0.4$ and $\alpha_r = \frac{1}{r+1}$.
- For Algorithm 1.4, we set $\vartheta_0 = 0.2$, $\epsilon = 0.85$, $f = \frac{1}{2}$ and $\delta_r = \frac{1}{1000r+1}$.
- For Algorithm 1.6, we set $\vartheta_1 = 0.6$, $\epsilon = 0.85$, $f = \frac{1}{2}$, $\alpha = 0.5$, $\varepsilon_r = \frac{1}{(1000r+1)^3}$ and $\delta_r = \frac{1}{1000r+1}$.
- For Algorithm 1, we set $\vartheta_1 = 1$, $\epsilon = 0.85$, $\rho_r = 1 + \frac{1}{r+1}$, $\alpha = 0.5$, $\varepsilon_r = \frac{1}{(1000r+1)^3}$, $\beta = 0.15$, $\varpi_r = \frac{1}{(1000r+1)^3}$, $f = \frac{1}{2}$, $\delta_r = \frac{1}{1000r+1}$ and $\sigma_r = \frac{1}{(5r+2)^2}$.

Also, we choose $t_{-1} = t_0 = (0, 0, \dots, 0) \in \mathbb{R}^{m\times n}$, $t_1 = (1, 1, \dots, 1) \in \mathbb{R}^{m\times n}$ and the regularization parameter of 10^{-6}, using MATLAB's motion blur operator with a length of 20 pixels and an angle of 45°. For comparison, we employed the standard test images of Broken arm X-ray (612×596) and Scoliosis X-ray (630×630). The algorithms terminate when the number of iterations reaches 500. The numerical results are reported in Table 2 and Figs. 2–5.

Remark 5.1. From the results in Figs. 2–5 and Table 2, it can be seen that the images obtained from the proposed Algorithm 1 have higher PSNR and SSIM values, indicating that the image quality recovered by the Algorithm 1 is better than the compared algorithms.

5.2. Data classification problem

In this classification, we apply a machine learning technique called the Extreme Learning Machine (ELM) proposed by Huang [24], which is structured as a single-hidden-layer feedforward networks (SLFNs). Let $\mathcal{K} := \{(x_r, y_r) : x_r \in \mathbb{R}^n, y_r \in \mathbb{R}^m, r = 1, 2, \dots, N\}$

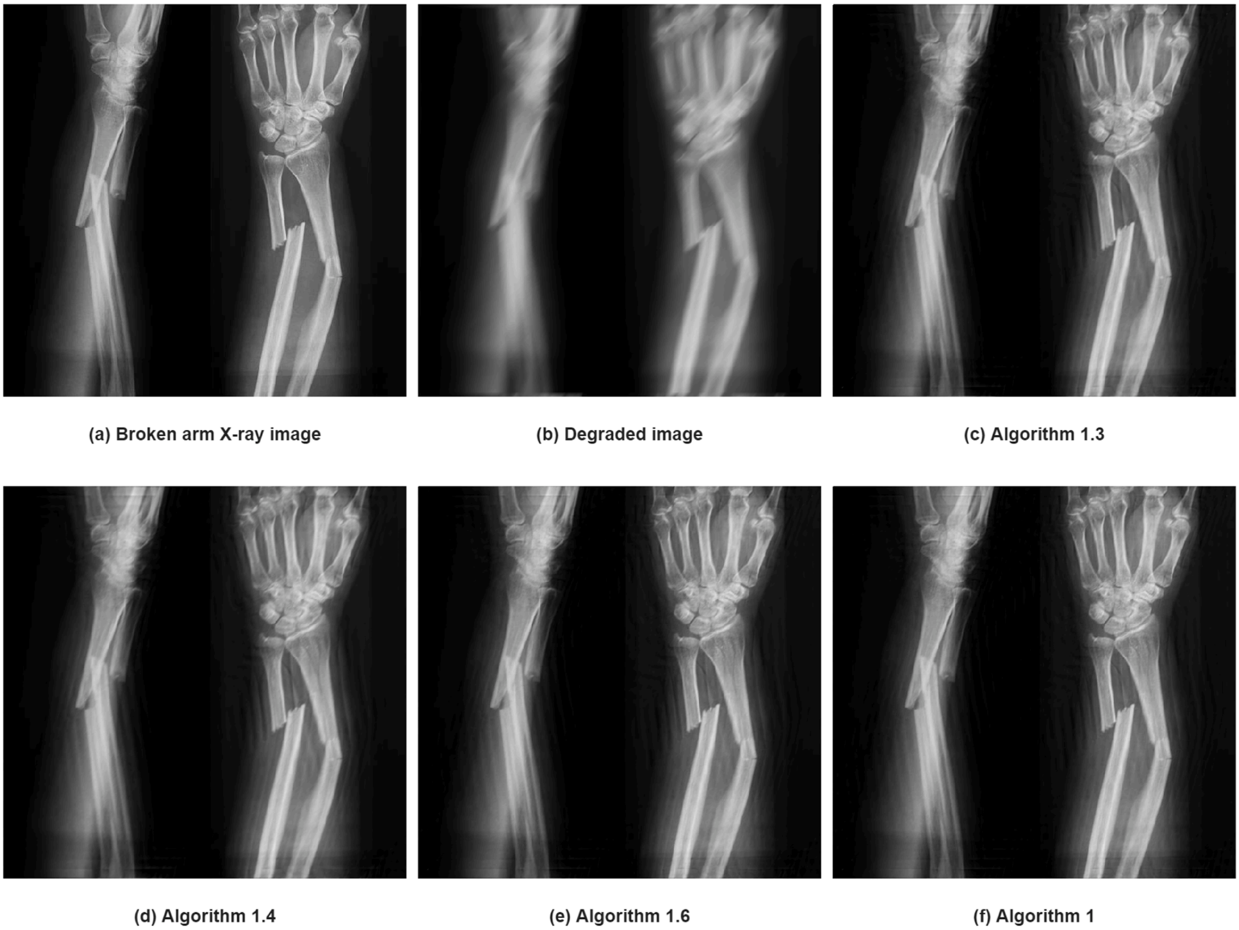

Fig. 2. The Broken arm X-ray image reconstructed by each algorithm.

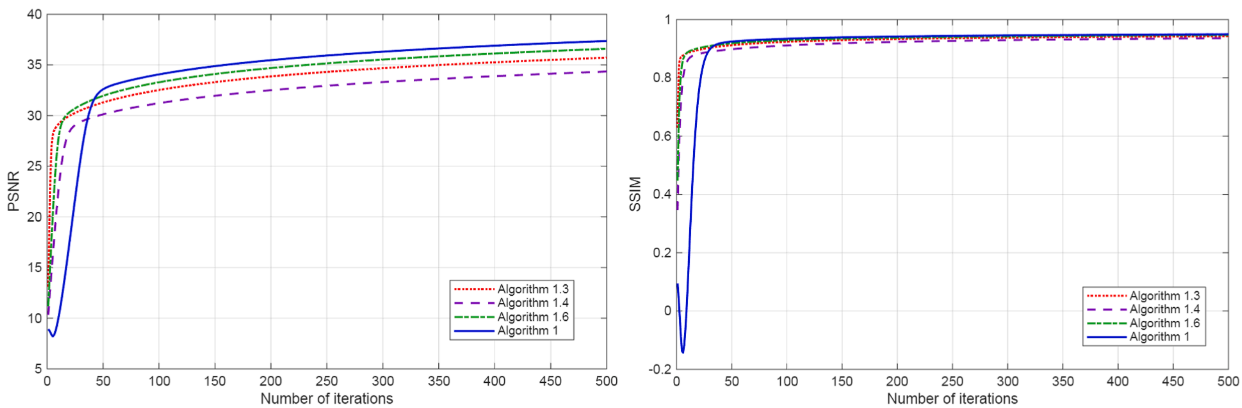

Fig. 3. The behavior of PSNR and SSIM for each algorithm of Broken arm X-ray.

be a training set of N samples, where x_r is input training data and y_r is the target. Given $\mathcal{M}$ hidden nodes, the output function of the ELM is defined as follows:

$$U_j = \sum_{i=1}^{\mathcal{M}} Q_i \frac{1}{1 + e^{-(\omega_i x_j + c_i)}},$$

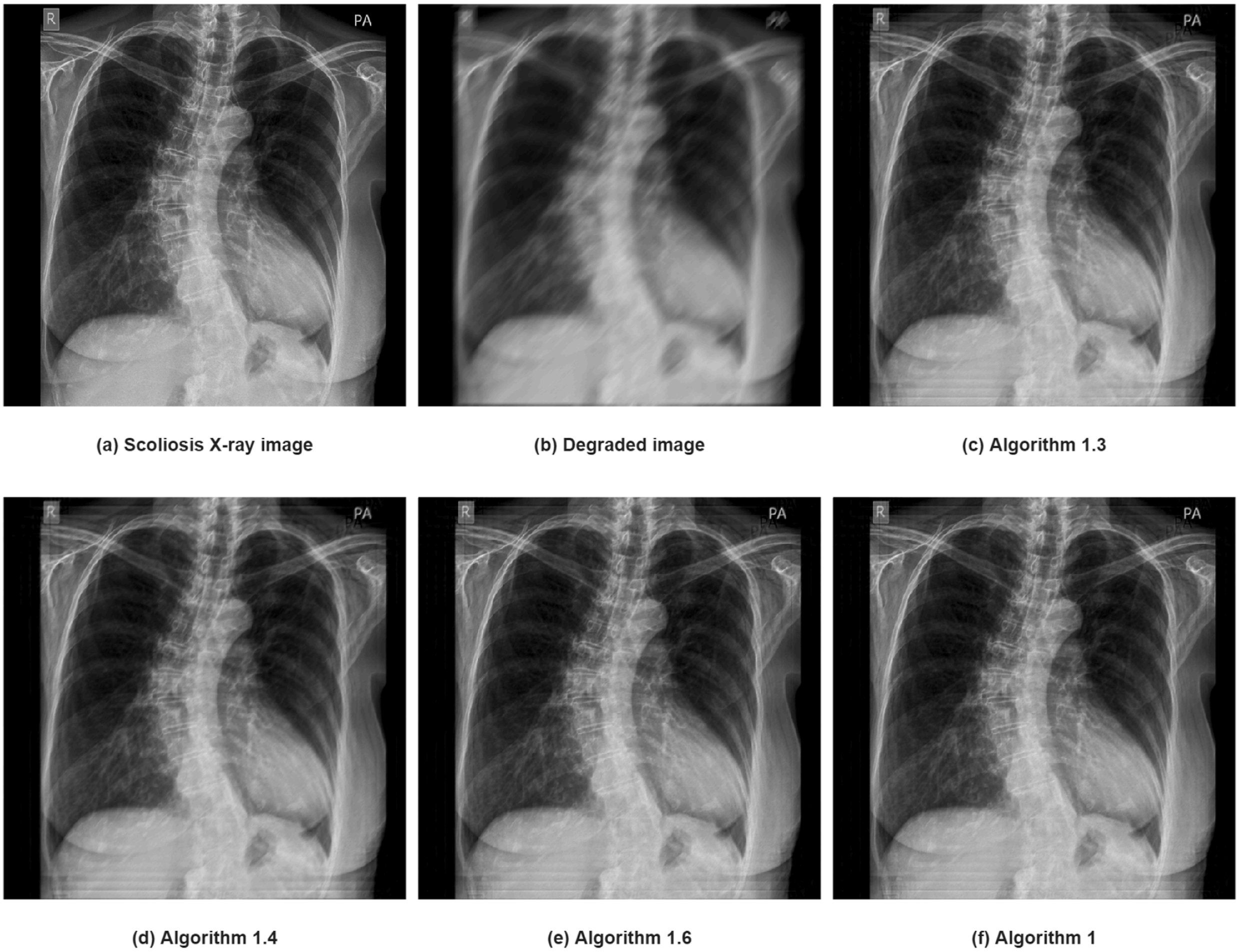

Fig. 4. The Scoliosis X-ray image reconstructed by each algorithm.

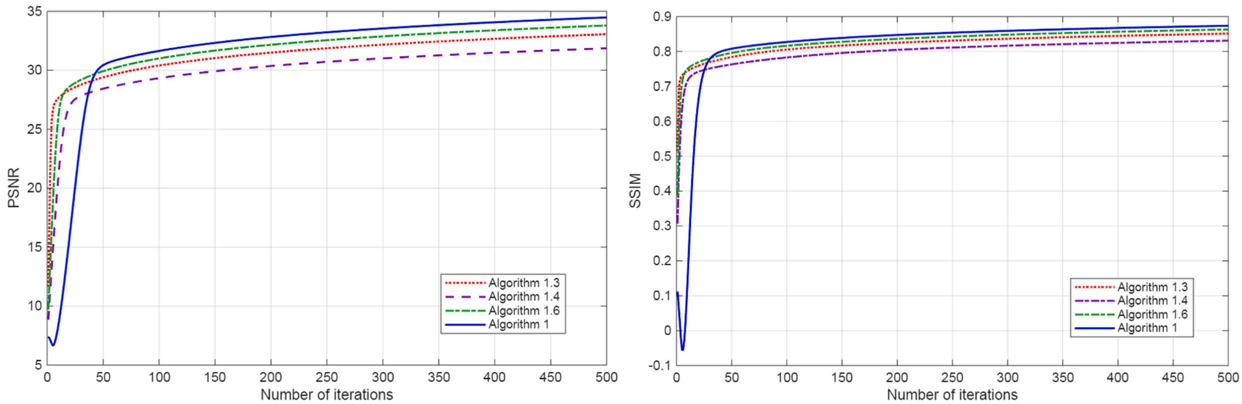

Fig. 5. The behavior of PSNR and SSIM for each algorithm of Scoliosis X-ray.

where ω_i is the weight between the input layer and the hidden layer and c_i is a bias vector for the hidden layer nodes. The goal is to find the optimal output weight Q_i. Thus, the output matrix of hidden layer B is generated as follows:

$$B = \begin{bmatrix} \frac{1}{1+e^{-(\omega_1 x_1 + c_1)}} & \cdots & \frac{1}{1+e^{-(\omega_{\mathcal{M}} x_1 + c_{\mathcal{M}})}} \\ \vdots & \ddots & \vdots \\ \frac{1}{1+e^{-(\omega_1 x_N + c_1)}} & \cdots & \frac{1}{1+e^{-(\omega_{\mathcal{M}} x_N + c_{\mathcal{M}})}} \end{bmatrix}$$

Table 2
Numerical performance obtained from different image experiments for all algorithms.

Algorithms	Broken arm X-ray		Scoliosis X-ray	
	PSNR	SSIM	PSNR	SSIM
Algorithm 1.3	35.7132	0.9432	33.0605	0.8513
Algorithm 1.4	34.3474	0.9362	31.8629	0.8310
Algorithm 1.6	36.5945	0.9469	33.7987	0.8631
Algorithm 1	37.3661	0.9498	34.4830	0.8738

The main goal of the ELM is to find

$$Q = [Q_1^T, \dots, Q_{\mathcal{M}}^T]^T \quad \text{such that} \quad BQ = \mathcal{T}, \tag{5.3}$$

where $\mathcal{T} = [t_1^T, \dots, t_N^T]^T$ is the target output matrix. These optimal weights can be computed as $Q = B^{\dagger}\mathcal{T}$, where $B^{\dagger}$ is the Moore-Penrose generalized inverse of B, though finding $B^{\dagger}$ may be challenging in practice. Therefore, obtaining a solution Q via convex minimization can help address this challenge. To address potential overfitting, Eq. (5.3) is reformulated as the following convex minimization problem:

$$\min_{Q \in \mathbb{R}^{\mathcal{M}}} \frac{1}{2}\|BQ - \mathcal{T}\|_2^2 + \tau\|Q\|_1, \tag{5.4}$$

where $\tau > 0$ is the regularized parameter. By applying Algorithm 1 to solve the problem (5.4), we are setting $X(Q) \equiv \nabla\left(\frac{1}{2}\|BQ - \mathcal{T}\|_2^2\right)$ and $Y(Q) \equiv \partial(\tau\|Q\|_1)$.

We also evaluated the performance of the classification algorithms using four evaluation metrics: precision, recall, F1-score and accuracy [25], which are defined as follows:

$$\begin{aligned} Precision &= \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100\% \\ Recall &= \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100\% \\ Accuracy &= \frac{True\ Positive + True\ Negative}{Total} \times 100\% \\ F1-score &= 2 \times \frac{Precision \times Recall}{Precision + Recall}. \end{aligned}$$

Next, we propose the formula for the cross-entropy loss function, which is used when there are more than two classes in a classification problem (multi-class classification) [26], and it can be calculated as follows:

$$Loss = -\sum_{i=1}^{G} \hat{\varphi}_i \log(\varphi_i),$$

where G is the number of classes, φ_i denotes the predicted probability for class ith and $\hat{\varphi}_i$ denotes the ith true class of training samples.

5.2.1. Osteoporosis Clinical Dataset

In this subsection, we present the applicability of the proposed method to data classification problem for osteoporosis patients from the publicly Harvard Dataverse, available on the online website [27]. The osteoporosis is defined as low bone mineral density resulting from altered bone microstructure, which ultimately predisposes patients to low-impact, fragility fractures. Osteoporotic fractures significantly reduce quality of life and are associated with increased morbidity, mortality, and disability. The risk of developing the condition is particularly high in postmenopausal women. To prevent bone loss and strengthen weak bones, a combination of medical management, a healthy diet, and maintenance of a healthy body weight is recommended. We used osteoporosis data consisting of 38 features that indicate characteristics across various age groups and 1463 instances, from young patients to the elderly.

Osteoporosis can be classified into four distinct stages. Each stage is determined by bone density, which serves as an indicator of the disease. The four stages of osteoporosis include:

- Stage 1: Low bone density without symptoms and bone density scores are above -1.
- Stage 2: Further bone loss and potential onset of symptoms and bone density scores are -1 to -2.5. This condition is not yet osteoporosis but are at risk of developing it. The medical term for this is osteopenia.
- Stage 3: Increased fracture risk and noticeable symptoms and bone density scores are -2.5 or lower.
- Stage 4: Severe bone loss, recurrent fractures that significant impact on daily life and bone density scores are below -2.5 with a history of one or more fractures.

For the comparison of each algorithm, we split the data are partitioned into a training set comprising 80% of the total data and a validation set comprising the remaining 20%. The parameter settings include using the sigmoid as the activation function, the number

Table 3
Parameter configuration details for each algorithm.

	Algorithm 1.3	Algorithm 1.4	Algorithm 1.6	Algorithm 1
$\vartheta_0 = \vartheta_1 = \frac{1}{2\|B\|^2}$	✓	✓	✓	✓
$\alpha_r = \frac{1}{r+1}$	✓	–	–	–
$\epsilon = 0.5$	–	✓	✓	✓
$f = \frac{1}{6}$	–	✓	✓	✓
$\delta_r = \frac{1}{5r+1}$	–	✓	✓	✓
$\alpha = 7.5$	–	–	✓	✓
$\beta = 4.5$	–	–	–	✓
$\varepsilon_r = \frac{1}{(5r+1)^3}$	–	–	✓	✓
$\varpi_r = \frac{1}{(5r+1)^3}$	–	–	–	✓
$\rho_r = 1 + \frac{1}{r+1}$	–	–	–	✓
$\sigma_r = \frac{1}{(1000r+1)^2}$	–	–	–	✓

Table 4
Comparative performance results of all algorithms after 800th iterations.

	CPU Time	Precision	Recall	F1-score	Accuracy Test
Algorithm 1.3	0.0105	84.34	76.92	80.46	80.94
Algorithm 1.4	0.0112	83.50	75.94	79.54	80.25
Algorithm 1.6	0.0134	83.87	76.08	79.79	80.37
Algorithm 1	0.0016	84.74	77.66	81.04	81.28

Table 5
Comparative performance results of all algorithms.

	Iter.	CPU Time	Precision	Recall	F1-score	Accuracy Test
Algorithm 1.3	837	0.0103	85.04	77.35	81.01	81.28
Algorithm 1.4	901	0.0108	83.98	76.43	80.03	80.59
Algorithm 1.6	999	0.0112	84.59	77.07	80.65	81.05
Algorithm 1	782	0.0098	85.39	77.56	81.28	81.39

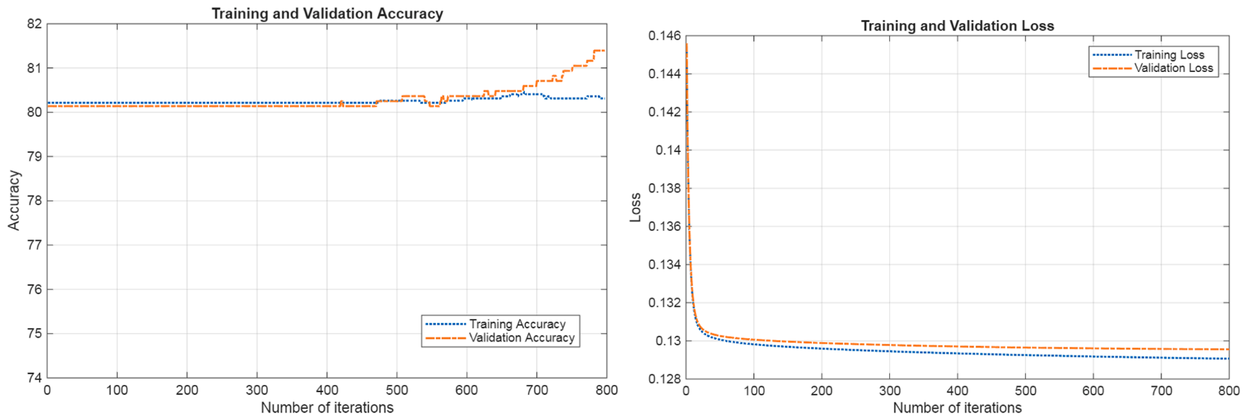

Fig. 6. Training and validation accuracy (left) and loss (right) graphs for Algorithm 1.

of hidden nodes $\mathcal{M} = 200$, and the regularization parameter $\tau = 10^{-6}$. In addition, the parameters for each algorithm are specified in Table 3.

The performance results of each algorithm at the 800th iterations are detailed in Table 4.

Next, we present the performance results of each algorithm shown in Table 5.

Remark 5.2. (i) From Table 4, the performance of each algorithm when using 800 iterations, it can be seen that Algorithm 1 has the highest accuracy at 81.28%, with the most notable precision, recall, and F1- score. It also demonstrates better CPU time performance. (ii) From Table 5, the results show that Algorithm 1 achieves an accuracy of 81.39%, with fewer number of iterations and better CPU time than the compared algorithms. It also demonstrates the best precision, recall, and F1- score.

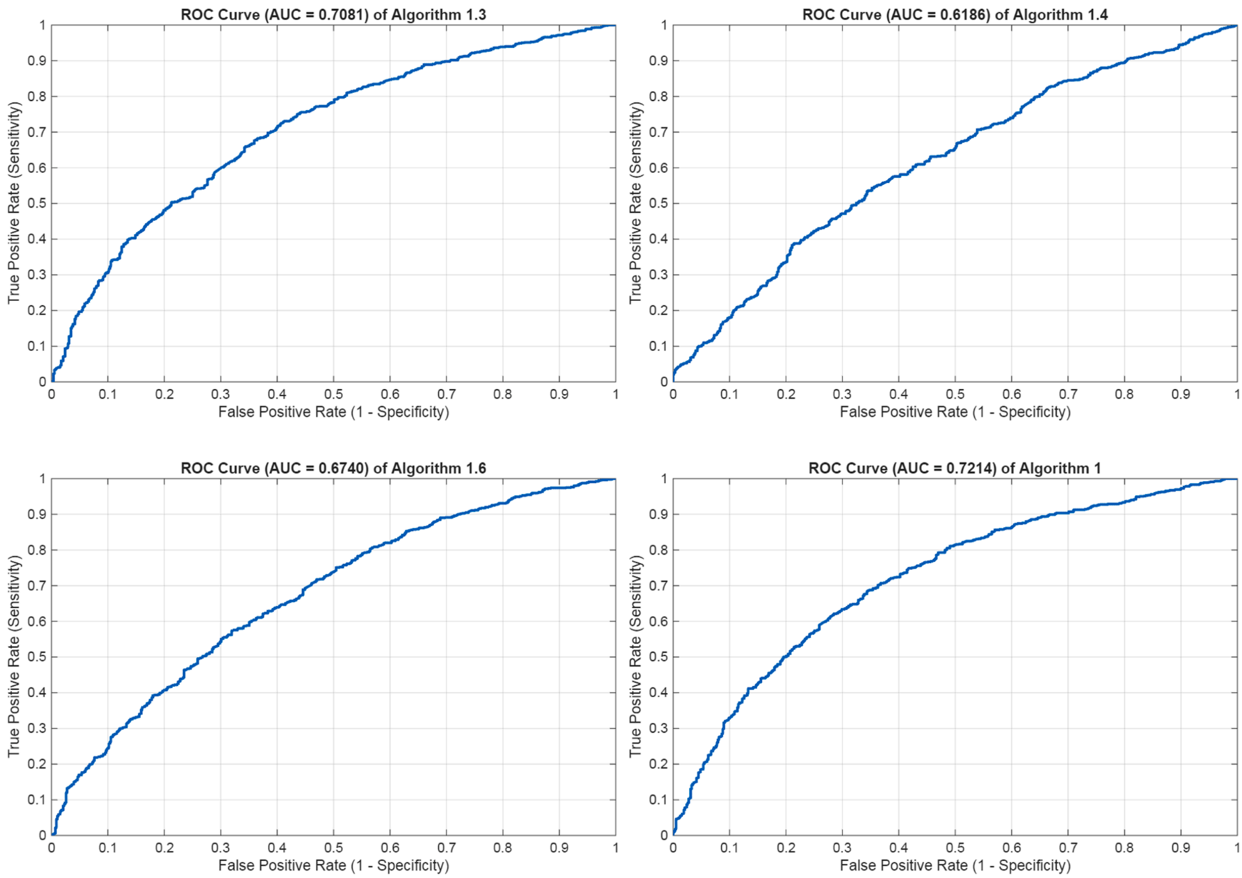

Fig. 7. Analysis of the ROC-AUC results for each algorithm.

Fig. 8. Representative knee radiographs illustrating bone density categories. From left to right: (Left) Normal, showing preserved trabecular pattern and cortical thickness; (Middle) Osteopenia, demonstrating mild reduction in bone density with partial trabecular thinning; and (Right) Osteoporosis, characterized by marked cortical thinning and decreased trabecular bone density.

For Algorithm 1, we present the trend graphs of training and validation accuracy and loss, the results of which are shown in Fig. 6. From Fig. 6, the accuracy and loss plots for Algorithm 1 indicate that the training and validation accuracy remaining closely aligned across iterations, suggesting minimal overfitting. The corresponding loss curves also display similar behavior, decreasing steadily over time and then stabilizing, which demonstrates that regularization and parameter selection effectively mitigate overfitting. This shows a performance trend that suggests that Algorithm 1 generalizes well to the validation data without overfitting the training set.

Next, we will display the ROC curve to demonstrate the model's classification performance, where a higher AUC value indicating a good performance of the model, as shown in Fig. 7.

From the graph shown in Fig. 7, the ROC-AUC metric is used to evaluate the performance of the osteoporosis classification model. The analysis shows that Algorithm 1 performs the best among all algorithms, achieving an AUC of 0.7214, which reflects its strong

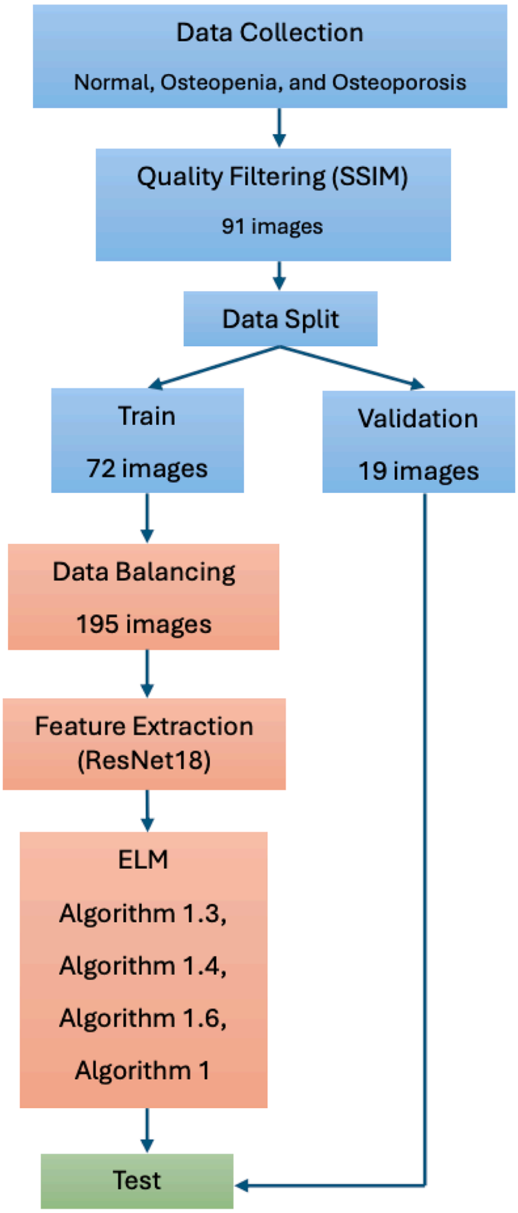

Fig. 9. Overview of the experimental workflow for multi-class knee osteoporosis classification.

ability to classify osteoporosis efficiently. Algorithms 1.3 and 1.6 achieve AUC values of 0.7081 and 0.6740, respectively, indicating reliable and robust classification performance. In contrast, Algorithm 1.4 performs less effectively, with an AUC of 0.6186. Therefore, Algorithm 1 demonstrates the best performance on the osteoporosis dataset, particularly in accurately predicting complex multi-class outcomes.

5.2.2. Multi-Class Knee Osteoporosis X-Ray Dataset

To evaluate the effectiveness of the proposed optimization-based learning framework, we employed the publicly available Multi-Class Knee Osteoporosis X-Ray Dataset obtained from Kaggle (Mohamed Gobara). This dataset consists of anterior–posterior (AP) knee radiographs categorized into three clinically relevant bone mineral density classes: Normal, Osteopenia, and Osteoporosis, see in Fig. 8. The dataset provides representative grayscale X-ray images reflecting progressive structural deterioration of bone tissue. The Normal class exhibits preserved trabecular architecture and cortical thickness, whereas the Osteopenia class shows mild trabecular thinning and early cortical reduction. The Osteoporosis class demonstrates marked cortical thinning and significant loss of trabecular density. These progressive morphological changes make the dataset suitable for multi-class image-based classification and automated screening research. Furthermore, this dataset has been widely utilized in recent studies on medical image classification and

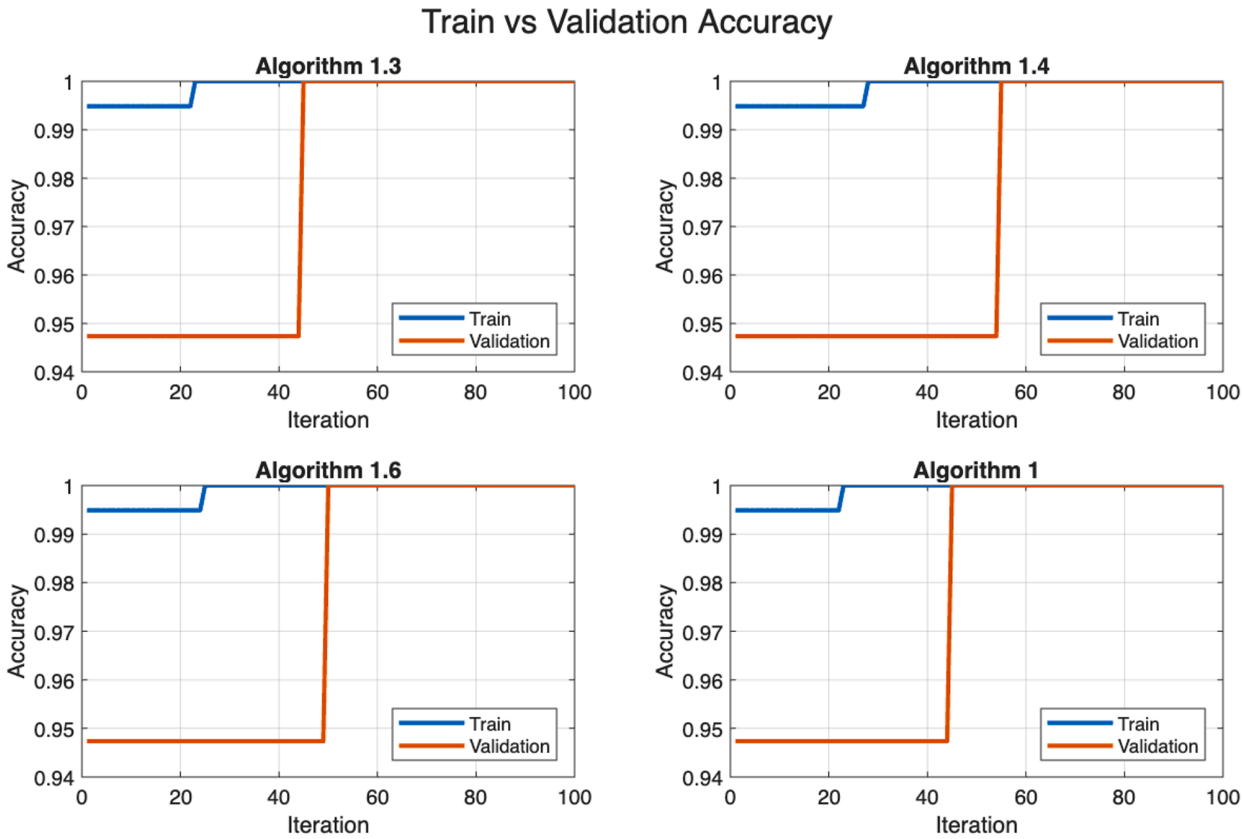

Fig. 10. Training and validation accuracy curves across 100 iterations for four optimization algorithms (Algorithms 1.3, 1.4, 1.6, and 1).

Table 6
Dataset distribution across processing stages.

Stage	Normal	Osteopenia	Osteoporosis	Total
SSIM-filtered dataset	8	81	2	91
Training set (raw)	6	65	1	72
Validation set	2	16	1	19
Balanced training set	65	65	65	195

osteoporosis detection, demonstrating its reliability and relevance for benchmarking deep learning and machine learning algorithms in clinical decision-support research [28,29]. In this study, the dataset is utilized to experimentally validate the proposed optimization algorithm integrated with the Extreme Learning Machine (ELM) framework. Specifically, we investigate the capability of the proposed algorithm to enhance feature learning and classification performance in a multi-class medical image screening task. The objective is to assess whether the optimized ELM model can effectively discriminate subtle structural differences in knee radiographs for early osteoporosis detection. This experimental setting enables systematic evaluation in terms of classification accuracy, convergence behavior, and computational efficiency, thereby demonstrating the applicability of the proposed approach in medical image–based decision support systems.

To systematically evaluate the effectiveness of the proposed optimization-based Extreme Learning Machine (ELM) framework for multi-class osteoporosis screening, a structured experimental workflow was designed, as illustrated in Fig. 9.

Fig. 9 shows that the process begins with data collection comprising Normal, Osteopenia, and Osteoporosis radiographs, followed by quality filtering using the Structural Similarity Index (SSIM), resulting in 91 eligible images. The dataset is then split into training (72 images) and validation (19 images) sets. To address class imbalance, data balancing is applied to the training set, increasing it to 195 images. Deep features are subsequently extracted using ResNet18 [30] and fed into Extreme Learning Machine (ELM)-based models, including Algorithm 1–1.6. The final performance evaluation is conducted on the test set.

Table 6 presents the distribution of knee X-ray images across different preprocessing stages, including SSIM filtering, data splitting, and class balancing. This stepwise summary highlights the class imbalance inherent in the original dataset and illustrates how the balancing strategy was applied exclusively to the training subset to ensure fair model learning.

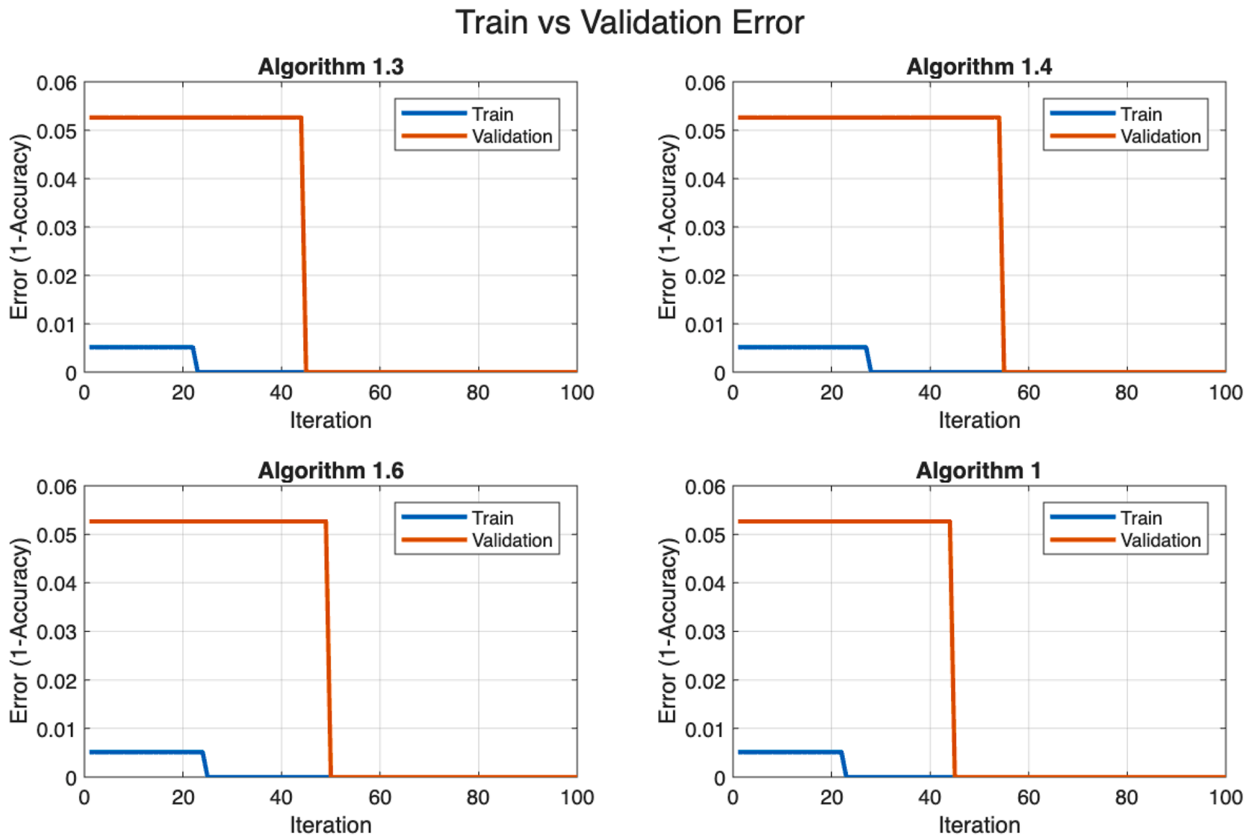

Fig. 11. Training and validation error (1-accuracy) versus iteration for the four optimization algorithms (Algorithms 1.3, 1.4, 1.6, and 1).

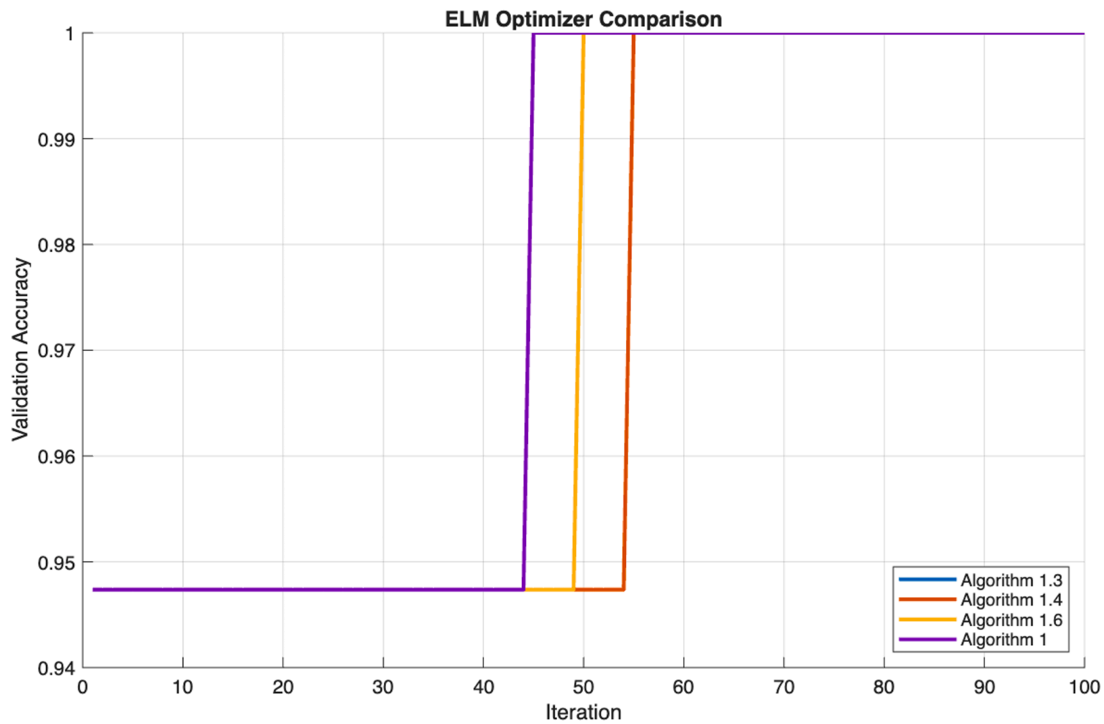

Fig. 12. Validation accuracy versus iteration for four ELM optimizers, showing convergence speed differences among the algorithms.

From the initial 91 SSIM-filtered images, the dataset is split into 72 training images and 19 validation images. A substantial class imbalance is observed, with Osteopenia dominating the dataset (81/91 images), while the Osteoporosis class contains only two samples. To address this imbalance and prevent biased learning, data balancing is applied to the training set, resulting in an equal distribution of 65 images per class (total 195 images). The validation set remains untouched to preserve unbiased performance evaluation.

Figs. 10 and 11 present the evolution of training and validation performance over 100 iterations for the four optimization variants integrated with the ELM classifier (Algorithms 1.3–1.6, and Algorithm 1). For a fair comparison with the competing algorithms, the parameter settings are chosen consistently with those reported in Table 3, with only the inertial parameters adjusted to $\alpha = 0.1$ and $\beta = 0.1$. This ensures that the observed performance differences are attributed to the structural characteristics of the algorithms rather than to inconsistent parameter tuning. Fig. 10 illustrates the accuracy curves, while Fig. 11 reports the corresponding error $(1 - \text{accuracy})$ trajectories. These curves provide insight into convergence behavior, stability, and generalization performance of each optimization strategy.

Across all algorithms, training accuracy rapidly approaches near-perfect performance within the first few dozen iterations, indicating efficient parameter updates and fast convergence of the optimization process. Validation accuracy initially remains moderate but exhibits a sharp improvement after approximately 40–60 iterations, eventually reaching perfect classification performance. The corresponding error curves consistently decrease to zero, confirming stable convergence without oscillatory behavior.

Fig. 12 presents a direct comparison of validation accuracy across 100 iterations for the four ELM-based optimization algorithms. This visualization highlights differences in convergence speed and stability, allowing clearer assessment of optimization efficiency beyond final accuracy alone.

All algorithms ultimately achieve perfect validation accuracy; however, noticeable differences are observed in their convergence rates. Algorithm 1 reaches optimal validation performance earlier than the other variants, indicating faster stabilization and more efficient parameter updating. Algorithms 1.6 and 1.4 follow closely, while Algorithm 1.3 converges slightly later. Despite these minor temporal differences, all methods demonstrate monotonic improvement without instability or oscillation.

6. Conclusions

In this article, we proposed a novel algorithm (Algorithm 1) for solving inclusion problems in real Hilbert spaces. The method is constructed by integrating a two-step inertial scheme with viscosity techniques, resulting in an enhanced iterative framework with improved stability and convergence behavior. Under appropriate assumptions, including monotonicity and Lipschitz continuity conditions, we established a strong convergence theorem for the generated sequence, thereby extending and strengthening several existing results in the literature. Numerical experiments further confirm the theoretical findings. The proposed algorithm consistently demonstrates faster convergence behavior and improved stability when compared with related methods under identical parameter settings. These results highlight the efficiency of the self-adaptive step-size mechanism and the contribution of the inertial–viscosity structure in accelerating the iterative process. In addition to the theoretical and numerical analysis, we validated the practical applicability of the proposed method in two real-world scenarios: medical image recovery and data classification. In image restoration tasks, the algorithm effectively enhances structural preservation and reconstruction quality. When embedded within an Extreme Learning Machine framework for multi-class medical data classification, the method exhibits strong generalization capability, stable convergence patterns, and competitive computational efficiency.

Author contributions

Suthep Suantai: Writing – Methodology, Formal analysis, funding acquisition; **Preeyanuch Chuasuk**: Writing – review & editing, Investigation, Conceptualization; **Watcharaporn Cholamjiak**: Writing – review & editing, Software; **Bing Tan**: Writing – review & editing, Investigation; **Pronpat Peeyada**: Writing – original draft, Methodology, Formal analysis, project administration, funding acquisition. All authors reviewed and confirmed the final manuscript before publication.

Ethics statement

This study used publicly available, de-identified data obtained from online sources and did not involve direct interaction with human participants or identifiable personal information. Ethical review was conducted, and the study was granted exemption from human research ethics approval by the Walailak University Ethics Committee (Approval No. WUEC-26-230-01).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work supported by Walailak University under the New Researcher Development scheme (Contract Number WU69202) and Chiang Mai University. W. Cholamjiak would like to thank University of Phayao and Thailand Science Research and Innovation Fund (Fundamental Fund 2026, Grant No. 2253/2568). P. Chuasuk would like to thank Burapha University.

Code availability

The MATLAB codes used in this study are available from the corresponding author upon reasonable request.

Data availability

The Broken arm X-ray image used in this research is publicly available at https://neuwirthlaw.com/broken-wrist-without-surgery-43000-vs-broken-wrist-with-surgery-100000/

The Scoliosis X-ray image used in this research is publicly available at https://radiopaedia.org/cases/scoliosis-6

The osteoporosis clinical dataset is publicly available at https://dataverse.harvard.edu/file.xhtml?fileId=6563909&version=1.0

The multi-class knee osteoporosis X-Ray dataset is publicly available at https://www.kaggle.com/datasets/mohamedgobara/multi-class-knee-osteoporosis-x-ray-dataset.

References

[1] T.O. Alakoya, O.T. Mewomo, Viscosity S-iteration method with inertial technique and self-adaptive step size for split variational inclusion, equilibrium and fixed point problems, Comput. Appl. Math. 41 (1) (2022) 39.
[2] A. Adamu, D. Kitkuan, T. Seangwattana, An accelerated halpern-type algorithm for solving variational inclusion problems with applications, Bangmod Int. J. Math. Comput. Sci. 8 (2022) 37–55.
[3] A. Adamu, P. Kumam, D. Kitkuan, A. Padcharoen, A Tseng-type algorithm for approximating zeros of monotone inclusion and J-fixed-point problems with applications, Fixed Point Theory Algorithms Sci. Eng. 2023 (1) (2023) 3.
[4] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. J. Issued Courant Inst. Math. Sci. 57 (11) (2004) 1413–1457.
[5] J. Duchi, Y. Singer, Efficient online and batch learning using forward backward splitting, J. Mach. Learn. Res. 10 (2009) 2899–2934.
[6] H. Raguet, J. Fadili, G. Peyré, A generalized forward-backward splitting, SIAM J. Imaging Sci. 6 (3) (2013) 1199–1226.
[7] A. Beck, M. Teboulle, Gradient-based algorithms with applications to signal-recovery problems, Found. Comput. Math. 10 (1) (2010) 73–100.
[8] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul. 4 (4) (2005) 1168–1200.
[9] R. Suparatulatorn, W. Liawrungrueang, T. Mouktonglang, W. Cholamjiak, An algorithm for variational inclusion problems including quasi-nonexpansive mappings with applications in osteoporosis prediction, AIMS Math. 10 (2) (2025) 2541–2561.
[10] M. Verma, D.R. Sahu, K.K. Shukla, VAGA: a novel viscosity-based accelerated gradient algorithm: convergence analysis and applications, Appl. Intell. 48 (9) (2018) 2613–2627.
[11] Y.H. Yao, R.Q. Xu, A. Adamu, Y. Shehu, A new numerically fast relaxed forward-backward-forward algorithm with applications to compressed sensing and image processing, J. Oper. Res. Soc. China 13 (2025) 1–27.
[12] A. Adamu, D. Kitkuan, P. Kumam, A. Padcharoen, T. Seangwattana, Approximation method for monotone inclusion problems in real banach spaces with applications, J. Inequalities Appl. 2022 (2022) 70.
[13] C. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, Inverse Probl. 20 (1) (2003) 103.
[14] A. Hanjing, S. Suantai, A fast image restoration algorithm based on a fixed point and optimization method, Mathematics 8 (3) (2020) 378.
[15] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (6) (1979) 964–979.
[16] D.A. Lorenz, T. Pock, An inertial forward-backward algorithm for monotone inclusions, J. Math. Imaging Vis. 51 (2) (2015) 311–325.
[17] A. Gibali, D.V. Thong, Tseng type methods for solving inclusion problems and its applications, Calcolo 55 (4) (2018) 49.
[18] N. Kaewyong, K. Sitthithakerngkiet, Modified Tseng's method with inertial viscosity type for solving inclusion problems and its application to image restoration problems, Mathematics 9 (10) (2021) 1104.
[19] W. Takahashi, Nonlinear Functional Analysis, 2000.
[20] H. Brezis, Opérateurs Maximaux Monotones et Semi-Groupes de Contractions Dans Les Espaces de Hilbert, Vol. 5, Elsevier, 1973.
[21] H.K. Xu, Iterative algorithms for nonlinear operators, J. Lond. Math. Soc. 66 (1) (2002) 240–256.
[22] P.E. Maingé, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, Set-Valued Anal. 16 (7) (2008) 899–912.
[23] H. Liu, J. Yang, Weak convergence of iterative methods for solving quasimonotone variational inequalities, Comput. Optim. Appl. 77 (2) (2020) 491–508.
[24] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[25] N. Nasir, P. Oswald, F. Barneih, O. Alshaltone, M. AlShabi, T. Bonny, A. Al Shammaa, Hypertension classification using machine learning part II, in: 2021 14th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2021, pp. 459–463.
[26] Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, Q. Wu, MPCE: a maximum probability based cross entropy loss function for neural network classification, IEEE Access 7 (2019) 146331–146341.
[27] Harvard Dataverse, Bone mineral density, Unknown, https://dataverse.harvard.edu/file.xhtml?fileId=6563909&version=1.
[28] N. Elreshidy, O. Gamal, A. Hosny, M. Deshish, M. Elarabie, A. Elnasharty, E. Elsharawy, Evaluating deep learning architectures for multi-class classification of medical X-ray images, in: 2025 1st Future International Conference on Artificial Intelligence and Cybersecurity (FICAC), IEEE, 2025, pp. 128–134.
[29] S.L. Mohanty, A.K. Pati, S. Saha, S. Gayen, A. Pati, A. Panigrahi, Enhanced deep learning for multi-class osteoporosis detection via feature fusion of diverse pre-trained models from knee X-ray images, in: 2025 International Conference on Responsible, Generative and Explainable AI (ResGenXAI), IEEE, 2025, pp. 1–6.
[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.