J. Math. Computer Sci., 38 (2025), 201–213

Online: ISSN 2008-949X

Journal of Mathematics and Computer Science

Journal Homepage: www.isr-publications.com/jmcs

A bi-inertial Mann projective forward-backward splitting algorithm for variational inclusion problems with application to lung cancer screening

Pronpat Peeyada[a], Watcharaporn Cholamjiak[a], Kanokwatt Shiangjen[b,*]

[a]School of Science, University of Phayao, Phayao 56000, Thailand.

[b]School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand.

Abstract

This paper proposes a bi-inertial Mann projective forward-backward splitting algorithm to solve the variational inclusion problem in real Hilbert spaces. We establish a weak convergence result under mild conditions commonly used in convergence analysis. Additionally, we present a series of numerical experiments to demonstrate the efficiency of our algorithm compared to existing methods. Finally, we applied our algorithm to classify data using the lung cancer dataset, achieving the highest testing accuracy of 90.32%, surpassing other documented algorithms. Our results indicate that the proposed algorithm offers a practical solution for detecting lung cancer.

Keywords: Forward-backward splitting algorithm, variational inclusion problem, extreme learning machine, data classification, lung cancer dataset.

2020 MSC: 46E20, 52A07, 68Q04.

1. Introduction

Throughout this article, we assume that H is a real Hilbert space. Let $\mathbb{A} : H \to H$ be a single-valued monotone mapping and $\mathbb{B} : H \to 2^H$ be a set-valued monotone mapping. A variational inclusion problem (VIP) involves finding a point $x^* \in H$ such that:

$$0 \in (\mathbb{A} + \mathbb{B})x^*. \tag{1.1}$$

The study of the VIP is central to applying mathematics, underpinning fundamental concepts such as convex minimization, split feasibility, fixed points, saddle points, variational inequalities, and equilibrium problems. It also serves as a versatile model with applications in signal processing, control system optimization, image reconstruction, statistical and machine learning, quantum mechanics, filtration theory, and more, as evidenced by the existing literature [1, 7, 14, 15, 18, 26, 28]. This broad applicability has

*Corresponding author

Email addresses: pronpat.pee@gmail.com (Pronpat Peeyada), watcharaporn.ch@up.ac.th (Watcharaporn Cholamjiak), kanokwatt.sh@up.ac.th (Kanokwatt Shiangjen)

doi: 10.22436/jmcs.038.02.04

Received: 2024-07-27 *Revised:* 2024-09-08 *Accepted:* 2024-10-23

attracted dedicated researchers who rigorously explore the problem and develop tailored algorithms for approximating solutions.

This paper focuses on developing fast and efficient iterative algorithms for solving the VIP. In recent years, scholarly interest in addressing the VIP has grown significantly, leading to numerous researchers' development of various methodologies [4, 9, 10, 25, 27, 29, 30] and their associated citations. We will begin by reviewing some established algorithms for solving the VIP in the literature before introducing our methods. One of the earliest methods for solving equation (1.1) is the forward-backward splitting method, which is defined as follows:

$$x_{n+1} = J_\lambda^{\mathbb{B}}(x_n - \lambda \mathbb{A} x_n), \ \forall n \in \mathbb{N}, \tag{1.2}$$

where $J_\lambda^{\mathbb{B}} = (I + \lambda \mathbb{B})^{-1}$ is the resolvent of the operator $\mathbb{B}$ and $\lambda \in (0, \frac{2}{L})$. The sequences generated by algorithm (1.2) demonstrate weak convergence towards a solution of (1.1) when $\mathbb{B}$ is $\frac{1}{L}$-inverse strongly monotone (or cocoercive). Another viable condition for (1.2) convergence is requiring similar strong monotonicity for $\mathbb{A} + \mathbb{B}$.

In 1964, Polyak introduced the inertial extrapolation technique, also referred to as the heavy ball method, to enhance the convergence speed of optimization algorithms [22]. This approach starts with initial points $x_0, x_1 \in H$ and a parameter $\lambda > 0$, generating the sequence through the following iteration:

$$x_{n+1} = x_n + \sigma_n(x_n - x_{n-1}) - \lambda \nabla f(x_n), \quad n \geqslant 1, \tag{1.3}$$

where $f : H \to H$ is differentiable, and $\sigma_n \in [0, 1)$ represents the extrapolation coefficient for the inertial step $\sigma_n(x_n - x_{n-1})$. The seminal contribution to weak convergence is the algorithm introduced by Lorenz and Pock [17]. They added the inertial terms to the forward-backward splitting algorithm. Their algorithm, called inertial forward-backward algorithm (IFBSA) as follows: for an arbitrary point $x_0, x_1 \in H$, define the sequence $\{x_n\}$ by

$$\begin{cases} w_n = x_n + \sigma_n(x_n - x_{n-1}), \\ x_{n+1} = J_{\lambda_n}^{\mathbb{B}}(w_n - \lambda_n \mathbb{A} w_n), n \geqslant 1, \end{cases} \tag{1.4}$$

where $\sigma_n \in [0, 1)$ is an extrapolation factor and λ_n is a step size parameter in positive real interval. They demonstrated that the iterative sequence generated by IFBSA weakly converges to a zero of the sum of two maximal monotone operators, $\mathbb{A}$ and $\mathbb{B}$.

Iyiola and Shehu [12] introduced the following two-point inertial proximal point algorithm (TPIPA) for approximating the solution of the monotone inclusion problem in Hilbert spaces:

$$\begin{cases} w_n = x_n + \sigma_n(x_n - x_{n-1}) + \beta_n(x_{n-1} - x_{n-2}), \\ x_{n+1} = (1 - \eta_n) w_n + \eta_n J_{\lambda_n}^{\mathbb{B}}(w_n), \end{cases} \tag{1.5}$$

where $\lambda_n > 0$, σ_n and β_n satisfy some conditions. Two-point/bi/double inertial techniques have been employed by many researchers to enhance the efficiency of algorithms [5, 13, 21].

Building on the work above, we introduce a bi-inertial proximal point algorithm that integrates the projection method with the forward-backward splitting algorithm. This proposed algorithm solves variational inclusion problems within real Hilbert spaces. We establish the weak convergence theorem for our iterative technique under certain mild assumptions. Additionally, we provide numerical examples to demonstrate the performance and effectiveness of our algorithm, discussing its implementation in infinite-dimensional Hilbert spaces. Finally, we apply our algorithm to a data classification problem related to predicting lung cancer. We calculate and compare key performance metrics to evaluate which classification model delivers more accurate predictions for the dataset. Our results indicate that the proposed algorithm performs better in handling classification problems. The structure of the paper is organized as follows. Section 1 provides the background, motivation, and key contributions of the research. Section 2 covers the fundamental definitions and necessary lemmas for the proofs. Section 3 describes the construction of the proposed algorithm, including convergence analysis and numerical examples. Finally, Section 4 discusses the application of the algorithm to classify breast cancer data, along with performance metrics for evaluation.

2. Preliminaries

Throughout this manuscript, let $\langle .,. \rangle$ and $\|.\|$ denote the inner product and norm on H, respectively. The weak and strong convergence will be denoted by "$\rightharpoonup$" and "$\to$", respectively. We will present several definitions and results that will be useful in proving our main theorem.

Definition 2.1 ([24]). Let $\mathbb{B} : H \to 2^H$ be a multivalued mapping. Then $\mathbb{B}$ is called

(i) monotone if $\langle r - s, x - w \rangle \geqslant 0$, $\forall (x, r), (w, s) \in \text{Graph}(\mathbb{B})$ (the graph of mapping $\mathbb{B}$),
(ii) maximal monotone if $\text{Graph}(\mathbb{B})$ is not properly contained in the graph of any other monotone mapping.

Definition 2.2 (Combettes2). A mapping $\mathbb{A} : H \to H$ is called

(i) L-Lipschitz continuous, if $\|\mathbb{A}x - \mathbb{A}w\| \leqslant L\|x - w\|$ for all $x, w \in H$, where L is a positive number;
(ii) nonexpansive, if $\|\mathbb{A}x - \mathbb{A}w\| \leqslant \|x - w\|$ for all $x, w \in H$;
(iii) firmly nonexpansive, if $\|\mathbb{A}x - \mathbb{A}w\|^2 \leqslant \langle x - w, \mathbb{A}x - \mathbb{A}w \rangle$ for all $x, w \in H$;
(iv) ξ-inverse strongly monotone if $\xi\mathbb{A}$ is firmly nonexpansive when $\xi > 0$.

It is well-known that every ξ-cocoercive mapping is $\frac{1}{\xi}$-Lipschitz continuous and monotone. It is well known that when $\mathbb{B} : H \to 2^H$ is a maximal monotone operator and $\lambda > 0$, the resolvent $J_\lambda^{\mathbb{B}}$ is single-valued and firmly nonexpansive [19]. The existence of the resolvent operator $J_\lambda^{\mathbb{B}}$ follows from the fact that $\mathbb{B}$ is maximal monotone, which, by Minty's theorem [20], ensures that $(I + \lambda\mathbb{B})$ is surjective. Moreover, using the monotonicity property of $\mathbb{B}$, the single-valuedness of $J_\lambda^{\mathbb{B}}$ can be established by contradiction.

Lemma 2.3 (Alvarez). *Let $\{\tau_n\}$, $\{\rho_n\}$, and $\{\kappa_n\}$ be the sequences in $[0, \infty)$, such that for each $n \geqslant 1$, $\tau_{n+1} \leqslant \tau_n + \rho_n(\tau_n - \tau_{n-1}) + \kappa_n$ and $\lim_{n\to+\infty} \kappa_n < +\infty$. Also, suppose that there exists a real number ρ with $0 \leqslant \rho_n \leqslant \rho < 1$, for all $n \geqslant 1$. Then, the following are true:*

(i) $\sum_{n=1}^{\infty} [\tau_n - \tau_{n-1}]_+ < +\infty$, *where* $[\kappa]_+ = \max\{\kappa, 0\}$;
(ii) $\lim_{n\to+\infty} \tau_n = \tau^*$, *for some* $\tau^* \in [0, \infty)$.

Lemma 2.4 (Lopez). *Let $\mathbb{A} : H \to H$ be a ξ-inverse strongly monotone mapping and $\mathbb{B} : H \to 2^H$ be a maximal monotone mapping. Then,*

(i) *for* $\lambda > 0$, $\text{Fix}(J_\lambda^{\mathbb{B}}(I - \lambda\mathbb{A})) = (\mathbb{A} + \mathbb{B})^{-1}(0)$;
(ii) *for* $0 < \lambda \leqslant \bar{\lambda}$ *and* $x \in H$, $\|x - J_\lambda^{\mathbb{B}}(I - \lambda\mathbb{A})x\| \leqslant 2\|x - J_{\bar{\lambda}}^{\mathbb{B}}(I - \bar{\lambda}\mathbb{A})x\|$.

Lemma 2.5 (Goebel). *Let H be a Hilbert space and let $\mathbb{A} : H \to H$ be a nonexpansive mapping with $\text{Fix}(\mathbb{A}) \neq \emptyset$, where $\text{Fix}(\mathbb{A})$ is the set of all fixed points of $\mathbb{A}$ that is $\text{Fix}(\mathbb{A}) = \{x^* \in H : x^* = \mathbb{A}x^*\}$. If, for any $\{x_n\}$ in H such that $x_n \rightharpoonup x$ and $\|x_n - \mathbb{A}x_n\| \to 0 \Rightarrow x \in \text{Fix}(\mathbb{A})$.*

Lemma 2.6 ([3]). *Let C be a nonempty subset of H. Also, let $\{x_n\}$ be a sequence which satisfies the following properties:*

(i) $\lim_{n\to\infty} \|x_n - x\|$ *exists for every* $x \in C$;
(ii) *every weak sequential cluster point of* $\{x_n\}$ *belongs to* C.

Then $\{x_n\}$ converges weakly to a point in C.

3. Main results

We begin by delineating the assumptions required in the proof of the main results in this section, with details of the first iterative scheme provided in Algorithm 1.

Algorithm 1 Bi-inertial Mann projective forward-backward algorithm (B-IMPFBA).

Initialization: Set $\{\sigma_n\}, \{\beta_n\} \subset [0,\infty)$, $\{\gamma_n\}, \{\eta_n\} \subset (0,1)$, and $\{\lambda_n\} \subset (0, 2\xi)$ and let $x_{-1}, x_0, x_1 \in H$.

Iterative Steps: Calculate the iterates x_{n+1} as follows.

Step 1. Calculate $w_n = x_n + \sigma_n(x_n - x_{n-1}) + \beta_n(x_{n-1} - x_{n-2})$.
Step 2. Calculate $z_n = w_n + \gamma_n(x_n - w_n)$.
Step 3. Calculate $x_{n+1} = P_C((1-\eta_n)J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})x_n + \eta_n J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n)$. Set $n := n+1$ and go to Step 1.

Assumption 3.1. *Consider the following conditions:*

(A1) $\mathbb{A} : H \to H$ *is* ξ*-inverse strongly monotone mapping;*
(A2) $\mathbb{B} : H \to 2^H$ *is set-valued maximal monotone mapping;*
(A3) C *is closed and convex nonempty subset in a real Hilbert space* H*;*
(A4) $\Omega := (\mathbb{A}+\mathbb{B})^{-1}(0) \cap C$ *is nonempty.*

Assumption 3.2. *Consider the following conditions:*

(C1) $\sum_{n=1}^{\infty} \sigma_n\|x_n - x_{n-1}\| < \infty$ *and* $\sum_{n=1}^{\infty} \beta_n\|x_{n-1} - x_{n-2}\| < \infty$*;*
(C2) $0 < \liminf_{n\to\infty} \lambda_n \leqslant \limsup_{n\to\infty} \lambda_n < 2\xi$*;*
(C3) $\limsup_{n\to\infty} \eta_n < 1$.

Theorem 3.3. *Let* $\{x_n\}$ *be generated by Algorithm 1 when Assumptions 3.1 and 3.2 hold. Then* $\{x_n\}$ *converges weakly to an element of* Ω.

Proof. Let $x^* \in \Omega$. Since $J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})$ and P_C are nonexpansive mappings, and $\mathbb{A}$ is ξ-inverse strongly monotone, we have

$$\begin{aligned}
\|x_{n+1} - x^*\| &= \|P_C((1-\eta_n)J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})x_n + \eta_n J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n) - x^*\| \\
&\leqslant (1-\eta_n)\|J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})x_n - x^*\| + \eta_n\|J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n - x^*\| \\
&\leqslant (1-\eta_n)\|x_n - x^*\| + \eta_n\|z_n - x^*\| \\
&\leqslant (1-\eta_n)\|x_n - x^*\| + \eta_n((1-\gamma_n)\|w_n - x^*\| + \gamma_n\|x_n - x^*\|) \\
&\leqslant (1-\eta_n)\|x_n - x^*\| + (1-\gamma_n)\eta_n(\|x_n - x^*\| + \sigma_n\|x_n - x_{n-1}\| \\
&\quad + \beta_n\|x_{n-1} - x_{n-1}\|) + \gamma_n\eta_n\|x_n - x^*\| \\
&= \|x_n - x^*\| + (1-\gamma_n)\eta_n(\sigma_n\|x_n - x_{n-1}\| + \beta_n\|x_{n-1} - x_{n-2}\|).
\end{aligned}$$

From Lemma 2.3 and Assumption 3.2 (C1), $\lim_{n\to\infty}\|x_n - x^*\|$ exists. This implies that both $\{x_n\}$ and $\{w_n\}$ are bounded. Since $J^{\mathbb{B}}_{\lambda_n}$ is a firmly nonexpansive mapping, we have

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 &= \|P_C((1-\eta_n)J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{B})x_n + \eta_n k J^{\mathbb{B}}_{\lambda_n}(I))\|^2 \\
&\leqslant (1-\eta_n)\|J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n)x_n - x^*\|^2 + \eta_n\|J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n - x^*\|^2 \\
&\leqslant (1-\eta_n)\|x_n - x^*\|^2 + \eta_n(\|z_n - x^*\|^2 - \lambda_n(2\xi-\lambda_n)\|\mathbb{A}z_n - \mathbb{A}x^*\|^2 \\
&\quad - \|(I-J^{\mathbb{B}}_{\lambda_n})(I-\lambda_n\mathbb{A})z_n - (I-J^{\mathbb{B}}_{\lambda_n})(I-\lambda_n\mathbb{A})x^*\|^2) \\
&= \|x_n - x^*\|^2 + 2(1-\gamma_n)\eta_n\sigma_n\langle\sigma_n(x_n - x_{n-1}) + \beta_n(x_{n-1} - x_{n-2}), w_n - x^*\rangle \\
&\quad - \eta_n\lambda_n(2\xi-\lambda_n)\|\mathbb{A}z_n - \mathbb{A}x^*\|^2 - \eta_n\|z_n - J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n - \lambda_n(\mathbb{A}z_n - \mathbb{A}x^*)\|^2.
\end{aligned}$$

This implies that

$$\begin{aligned}
&\eta_n\lambda_n(2\xi-\lambda_n)\|\mathbb{A}z_n - \mathbb{A}x^*\|^2 + \eta_n\|z_n - J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n - \lambda_n(\mathbb{A}z_n - \mathbb{A}x^*)\|^2 \\
&\quad \leqslant \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 + 2(1-\gamma_n)\eta_n\sigma_n\langle\sigma_n(x_n - x_{n-1}) + \beta_n(x_{n-1} - x_{n-2}), w_n - x^*\rangle.
\end{aligned} \tag{3.1}$$

Since $\lim_{n\to\infty}\|x_n - x^*\|$ exists, if follows from (3.1) and Assumption 3.2 that

$$\lim_{n\to\infty}\|\mathbb{A}z_n - \mathbb{A}x^*\| = \lim_{n\to\infty}\|z_n - J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n - \lambda_n(\mathbb{A}z_n - \mathbb{A}x^*)\|^2 = 0.$$

This implies that $\lim_{n\to\infty}\|z_n - J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n\| = 0$. Since $\liminf_{n\to\infty}\lambda_n > 0$, there exists $\lambda > 0$ such that $0 < \lambda \leqslant \lambda_n$. By Lemma 2.4 (ii), we obtain

$$\lim_{n\to\infty}\|z_n - J^{\mathbb{B}}_{\lambda}(I-\lambda\mathbb{A})z_n\| \leqslant \lim_{n\to\infty}\|z_n - J^{\mathbb{B}}_{\lambda_n}(I-\lambda_n\mathbb{A})z_n\| = 0. \tag{3.2}$$

By the assumption (C1), we have

$$\lim_{n\to\infty}\|x_n - w_n\| = \lim_{n\to\infty}\left(\sigma_n\|x_n - x_{n-1}\| + \beta_n\|x_{n-1} - x_{n-2}\|\right) = 0.$$

This implies that

$$\lim_{n\to\infty}\|x_n - z_n\| = \lim_{n\to\infty}(1-\gamma_n)\|x_n - w_n\| = 0. \tag{3.3}$$

We next let $\tilde{x}$ be a weak sequential cluster point of $\{x_n\}$. From (3.3), we obtain $\tilde{x}$ is also a weak sequential cluster point of $\{z_n\}$. Since $\{x_n\}$ is the sequence generated in C and C is closed convex set, $\tilde{x} \in C$. It follows from (3.2) that $\tilde{x} \in \Omega$ by Lemma 2.5. Since $\lim_{n\to\infty}\|x_n - \tilde{x}\|$ exists, by Opial's lemma (Lemma 2.6), we can conclude that $\{x_n\}$ converges weakly to $\tilde{x}$. Theorem 3.3 is completed. □

4. Numerical examples

In this section, we present some numerical experiments occurring in infinite-dimensional spaces to illustrate the performance of our algorithm.

Example 4.1. Let $H = L_2([0,1])$ with norm $\|x\| := \sqrt{\int_0^1 |x(t)|^2 dt}$, $\forall x \in L_2([0,1])$. We consider the following nonempty closed convex subset C in $L_2([0,1])$:

$$C = \left\{x \in L_2([0,1]) | \int_0^1 x(t)dt \leqslant \infty\right\}.$$

Define the mapping $\mathbb{A} : L_2([0,1]) \to L_2([0,1])$ and $\mathbb{B} : L_2([0,1]) \to L_2([0,1])$ by $\mathbb{A}x(t) := \frac{x(t)}{2}, \forall x \in L_2([0,1])$ and $\mathbb{B}x(t) := 5x(t), \forall x \in L_2([0,1])$. For the parameters, we choose $\sigma_n = 0.8$, $\beta_n = 0.4$, $\lambda_n = 0.25$, $\gamma_n = \frac{n}{10n+1}$, and $\eta_n = \frac{n}{2n+1}$. Select the starting points $x_{-1}, x_0, x_0 \in L_2([0,1])$ randomly. We select various initial values as follows:

Case I: $x_{-1} = t$, $x_0 = e^t$, $x_1 = t^2 - 3$;
Case II: $x_{-1} = t^3 - 3$, $x_0 = e^t + t$, $x_1 = 2t$;
Case III: $x_{-1} = \frac{2t^2 - e^{-t}}{2}$, $x_0 = \frac{e^t}{2}$, $x_1 = \frac{t^2}{2}$;
Case IV: $x_{-1} = e^t$, $x_0 = 2t^2$, $x_1 = t^2 + t + 1$.

We set the stopping criterion as $\|x_n - x_{n-1}\|^2 < 10^{-3}$ and perform computation experiments for Algorithm 1, IFBSA(1.4), and TPIPA(1.5). The numerical results are reported in Table 1 and Figure 1.

Table 1: Numerical computational results of Algorithm 1, IFBSA(1.4), and TPIPA(1.5).

		Algorithm 1	IFBSA (1.4)	TPIPA (1.5)
Case I	CPU time (sec)	0.2881	0.1707	0.9616
	No. of Iter.	6	9	36
Case II	CPU time (sec)	3.4682	1.5315	11.9877
	No. of Iter.	5	7	28
Case III	CPU time (sec)	3.3573	1.4527	5.8850
	No. of Iter.	4	6	12
Case IV	CPU time (sec)	2.7722	0.1744	5.7771
	No. of Iter.	6	8	22

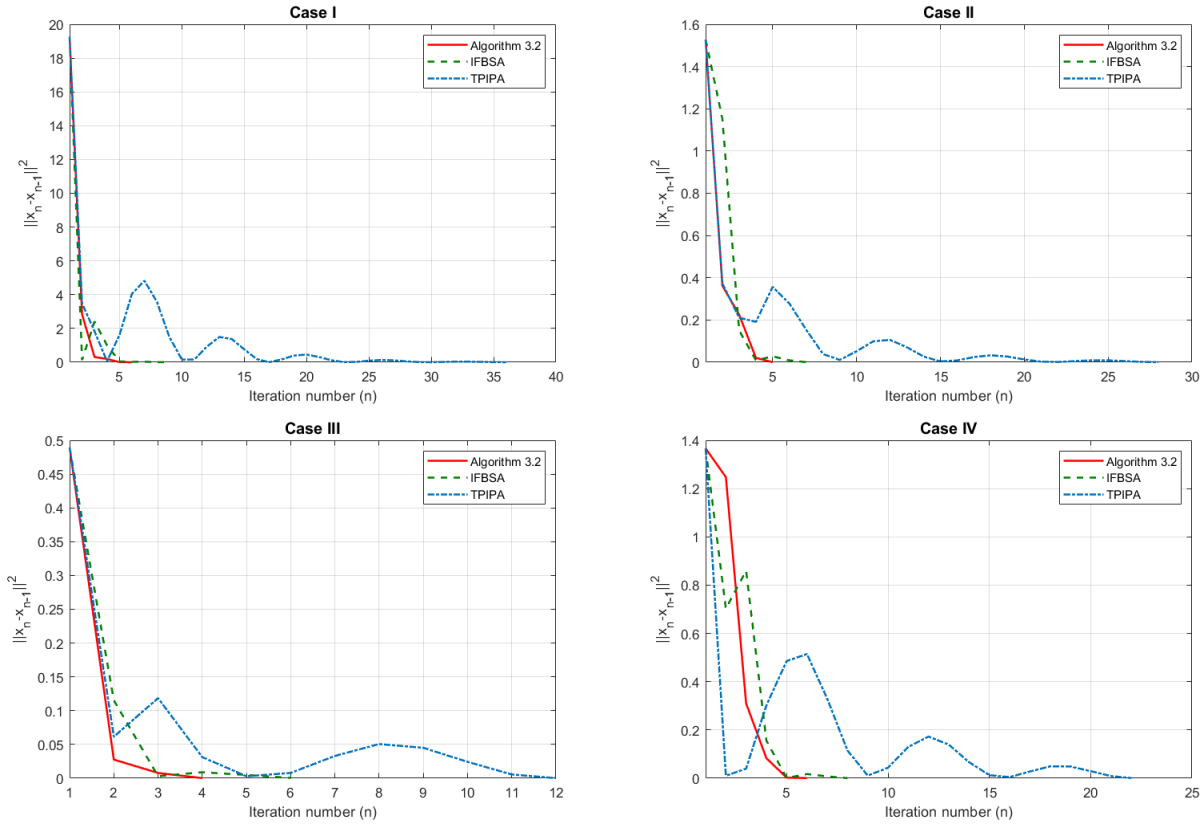

Figure 1: Numerical computational behavior of Algorithm 1, IFBSA (1.4), and TPIPA (1.5) for all four cases.

Based on Table 1 and Figure 1, it is evident that our algorithm performs better than IFBSA (1.4) and TPIPA (1.5) in terms of the number of iteration steps.

5. Numerical implementations in data classifications

Currently, many countries worldwide are grappling with issues related to particulate matter, particularly PM 2.5, an air pollutant adversely impacting the health of numerous individuals. Additionally, it hampers visibility. These smoke or dust particles often result from human activities like burning soil, garbage disposal through burning, and the combustion of fossil fuels. Furthermore, they can occur naturally during events such as sandstorms, dust storms, forest fires, and even the eruption of active volcanoes.

The global air pollution crisis, escalating in cities, industrial areas, and other regions, has prompted the World Health Organization (WHO) to warn. The WHO states that exposure to PM 2.5, or dust particles

no more significant than 2.5 micrometers in size, not only increases the likelihood of these particles being absorbed into the bloodstream and affecting bodily functions but can also lead to lung cancer when accumulated over an extended period. Information published in the medical journal 'Oncology Letter (PMC5920433) further reveals that PM 2.5 induces mutations and abnormal cell division in the body, creating conducive conditions for the onset of cancer. It was also found that PM 2.5 levels were measured at 22 micrograms per cubic meter, roughly equivalent to the exposure of smoking one cigarette. It is important to emphasize that the risk of developing lung cancer due to PM 2.5 exposure is not as high as that of smoking. Nonetheless, prolonged and extensive exposure to PM 2.5 can increase the risk of developing lung cancer.

Lung cancer may seem distant to those who don't smoke. However, according to the latest data from the World Health Organization (WHO), lung cancer ranks among the top eight causes of premature death and accounts for the highest number of cancer-related deaths worldwide. Lung cancer has become a silent killer, claiming many lives, primarily because it can develop and spread rapidly without manifesting any symptoms until the cancer cells have already spread extensively. It is only at this stage that patients can observe symptoms. Although we are aware of the causes and risk factors associated with lung cancer, only a tiny number of patients can detect it in its early stages.

In this study, we utilized lung cancer data obtained from Kaggle for classification. The dataset comprised 309 samples with 16 attributes. All the data were categorized into two classes: diseased and non-diseased. A total of 270 samples exhibited the disease, while 39 samples were free from the disease. Lung cancer information for each attribute is detailed in Table 2.

Table 2: Attributes information of Lung cancer dataset.

Attribute name	**Definitions and encoding**
Input	
Gender	1 := Male, 2 := Female
Age	Age of the patient
Smoking	1 := No, 2 := Yes
Yellow fingers	1 := No, 2 := Yes
Anxiety	1 := No, 2 := Yes
Peer pressure	1 := No, 2 := Yes
Chronic disease	1 := No, 2 := Yes
Fatigue	1 := No, 2 := Yes
Allergy	1 := No, 2 := Yes
Wheezing	1 := No, 2 := Yes
Alcohol	1 := No, 2 := Yes
Coughing	1 := No, 2 := Yes
Shortness of breath	1 := No, 2 := Yes
Swallowing difficulty	1 := No, 2 := Yes
Chest pain	1 := No, 2 := Yes
Output	
Lung cancer	1 := No, 2 := Yes

In this section, our focus is on the extreme learning machine (ELM) proposed by Huang et al. [11], which we apply to our algorithms in the machine learning data classification process. Let $\Upsilon := \{(x_n, t_n) : x_n \in \mathbb{R}^p, t_n \in \mathbb{R}^q, n = 1, 2, ..., N, p, q \in \mathbb{N}\}$ be a training set of N distinct samples, where x_n is an input training data and t_n is an input target. The output function of ELM for single-hidden layer feedforward neural networks (SLFNs) with M hidden nodes is defined as follows:

$$O_j = \sum_{i=1}^{M} \omega_i \frac{1}{1 + e^{-(v_i x_j + b_i)}},$$

where ω_i is the optimal output weight at the i^{th} hidden node, v_i is parameter weight, and b_i is the bias. The hidden layer output matrix $\mathbb{H}$ is generated as follows:

$$\mathbb{H} = \begin{bmatrix} \frac{1}{1+e^{-(v_1 x_1 + b_1)}} & \cdots & \frac{1}{1+e^{-(v_M x_1 + b_M)}} \\ \vdots & \ddots & \vdots \\ \frac{1}{1+e^{-(v_1 x_N + b_1)}} & \cdots & \frac{1}{1+e^{-(v_M x_N + b_M)}} \end{bmatrix}.$$

In experiments on regression and classification problem, the main goal of ELM is to find

$$\omega = [\omega_1{}^T, \omega_2{}^T, \ldots, \omega_M{}^T]^T \text{ such that } \mathbb{H}\omega = \mathbb{T}, \tag{5.1}$$

where $\mathbb{T} = [t_1{}^T, t_2{}^T, \ldots, t_N{}^T]^T$ is the training target data. To avoid overfitting in the model, problem (5.1) can be formulated as a convex minimization problem considered on a closed convex subset. We will consider four problem models as follows.

(i) Regularized least squares problem by L_1 (RLSP-L_1) or well-known called the least absolute shrinkage and selection operator (LASSO): for $\lambda > 0$,

$$\min_{\omega \in \mathbb{R}^M} \frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2 + \lambda\|\omega\|_1. \tag{5.2}$$

(ii) Regularized least squares problem by L_2 (RLSP-L_2): for $\lambda > 0$,

$$\min_{\omega \in \mathbb{R}^M} \frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2 + \lambda\|\omega\|_2^2. \tag{5.3}$$

(iii) Regularized least squares problem by L_1 with constrained within convex set L_1 (RLSP-L_1-CL_1): for $\lambda, \rho > 0$,

$$\min_{\omega \in C} \frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2 + \lambda\|\omega\|_1, \tag{5.4}$$

where $C = \{\omega : \|\omega\|_1 \leqslant \rho\}$.

(iv) Regularized least squares problem by L_2 with constrained within convex set L_2 (RLSP-L_2-CL_2): for $\lambda, \rho > 0$,

$$\min_{\omega \in C} \frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2 + \lambda\|\omega\|_2^2, \tag{5.5}$$

where $C = \{\omega : \|\omega\|_2^2 \leqslant \rho\}$.

For applying our algorithms to solve all of the convex minimization problems as above, we set our operator as in Table 3.

Table 3: Setting operators of our algorithms to solve all of the convex minimization problems (5.2)-(5.5).

Problems	**Setting operator of our algorithms**
RLSP-L_1	$\mathbb{A}(\omega) \equiv \nabla(\frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2)$, $\mathbb{B}(\omega) \equiv \partial(\lambda\|\omega\|_1)$, $C = H$
RLSP-L_2	$\mathbb{A}(\omega) \equiv \nabla(\frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2)$, $\mathbb{B}(\omega) \equiv \partial(\lambda\|\omega\|_2^2)$, $C = H$
RLSP-L_1-CL_1	$\mathbb{A}(\omega) \equiv \nabla(\frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2)$, $\mathbb{B}(\omega) \equiv \partial(\lambda\|\omega\|_1)$, $C = \{\omega : \|\omega\|_1 \leqslant \rho\}$
RLSP-L_2-CL_2	$\mathbb{A}(\omega) \equiv \nabla(\frac{1}{2}\|\mathbb{H}\omega - \mathbb{T}\|_2^2)$, $\mathbb{B}(\omega) \equiv \partial(\lambda\|\omega\|_2^2)$, $C = \{\omega : \|\omega\|_2^2 \leqslant \rho\}$

Evaluation metrics are essential tools for evaluating the performance of well-trained models on unseen test data. In examining all ensemble models, we conducted thorough tests using the designated test

dataset to measure accuracy, precision, recall, and F1-scores. The following metrics described in previous papers are referenced in [23]:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \times 100\%, \qquad \text{Precision} = \frac{T_P}{T_P + F_P} \times 100\%,$$
$$\text{Recall} = \frac{T_P}{T_P + F_N} \times 100\%, \qquad \text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})},$$

where T_P is true positive, F_P is false positive, T_N is true negative, and F_N is false negative.

The cross-entropy loss function is widely used in classification tasks for its quantitative assessment of how accurately a machine learning model classifies a dataset, particularly for binary classification problems. So the formula (5.6) is obtained:

$$\text{Loss} = -\frac{1}{S}\sum_{i=1}^{S} \mu_i \log \hat{\mu}_i + (1-\mu_i)\log(1-\hat{\mu}_i), \tag{5.6}$$

where S is the number of samples, μ_i is real value, and $\hat{\mu}_i$ is predicted value. For suitable parameters of our algorithm for each model problem in Table 3, we set $\lambda_n = \frac{0.3}{\|\mathbb{H}\|^2}$ and

$$\sigma_n = \begin{cases} \frac{\bar{\sigma}_n}{n^2\|x_n - x_{n-1}\|}, & \text{if } x_n \neq x_{n-1} \text{ and } n > N, \\ \bar{\sigma}_n, & \text{otherwise,} \end{cases} \qquad \beta_n = \begin{cases} \frac{\bar{\beta}_n}{n^2\|x_{n-1} - x_{n-2}\|}, & \text{if } x_{n-1} \neq x_{n-2} \text{ and } n > N, \\ \bar{\beta}_n, & \text{otherwise,} \end{cases}$$

$$\gamma_n = \begin{cases} \bar{\gamma}_n, & \text{if } n \leqslant N, \\ 0.9, & \text{otherwise,} \end{cases} \qquad \eta_n = \begin{cases} \bar{\eta}_n, & \text{if } n \leqslant N, \\ 0.5, & \text{otherwise,} \end{cases}$$

where N is the number of iterations at which we aim to stop. The essential parameters used in Algorithm 1, IFBSA (1.4), and TPIPA (1.5) algorithms can be seen in Table 4.

Table 4: Chosen parameters of each algorithm.

Algorithm	$\bar{\sigma}_n$	$\bar{\beta}_n$	$\bar{\gamma}$	$\bar{\eta}_n$	λ	ρ
IFBSA (1.4)	$\frac{1}{n+1}$	-	-	-	10^{-3}	-
TPIPA (1.5)	$\frac{2^{12}}{\|x_n - x_{n-1}\|^5 + n^5 + 2^{12}}$	$\frac{10^{15}}{\|x_n - x_{n-1}\|^3 + n^3 + 10^{15}}$	-	$\frac{n}{2n+1}$	10^{-5}	-
Algorithm 1 (RLSP-L_1)	$\frac{1}{\|x_n - x_{n-1}\|^3 + n^3 + 2^{15}}$	$\frac{1}{100n+1}$	0.7	0.45	10^{-5}	-
Algorithm 1 (RLSP-L_2)	$\frac{2^{15}}{\|x_n - x_{n-1}\|^3 + n^3 + 2^{15}}$	$\frac{2^{10}}{\|x_n - x_{n-1}\|^2 + n^2 + 2^{10}}$	$\frac{n}{2n+1}$	$\frac{0.9n}{n+1}$	10^{-7}	-
Algorithm 1 (RLSP-L_1-CL_1)	$\frac{2^{12}}{\|x_n - x_{n-1}\|^5 + n^5 + 2^{12}}$	$\frac{2^{15}}{\|x_n - x_{n-1}\|^2 + n^2 + 2^{15}}$	0.9	0.5	10^{-7}	9
Algorithm 1 (RLSP-L_2-CL_2)	$\frac{1}{100n+1}$	$\frac{1}{2n+1}$	0.3	0.5	10^{-7}	4

Next, we set $\hat{A}$ be sigmoid function and let hidden node $M = 100$. We divided the dataset using the percentage split technique, allocating 80% of the data to the training set and 20% to the testing set. The following table shows the numerical training results when the model stops at the highest of stable accuracy.

Table 5 demonstrates that the algorithms under study exhibit superior performance in terms of accuracy, precision, recall, and F1-score, leading to the most accurate classification of lung cancer compared to the other algorithms evaluated.

Table 5: Comparison of the performance metrics of various algorithms in ELM.

Algorithm	Training time	Precision	Recall	F1-score	Accuracy
IFBSA (1.4)	0.0337	87.10	100.00	93.10	87.10
TPIPA (1.5)	0.0053	87.10	100.00	93.10	87.10
Algorithm 1 (RLSP-L_1)	0.0017	90.00	100.00	94.74	90.32
Algorithm 1 (RLSP-L_2)	0.0021	90.00	100.00	94.74	90.32
Algorithm 1 (RLSP-L_1-CL_1)	0.0020	90.00	100.00	94.74	90.32
Algorithm 1 (RLSP-L_2-CL_2)	0.0016	90.00	100.00	94.74	90.32

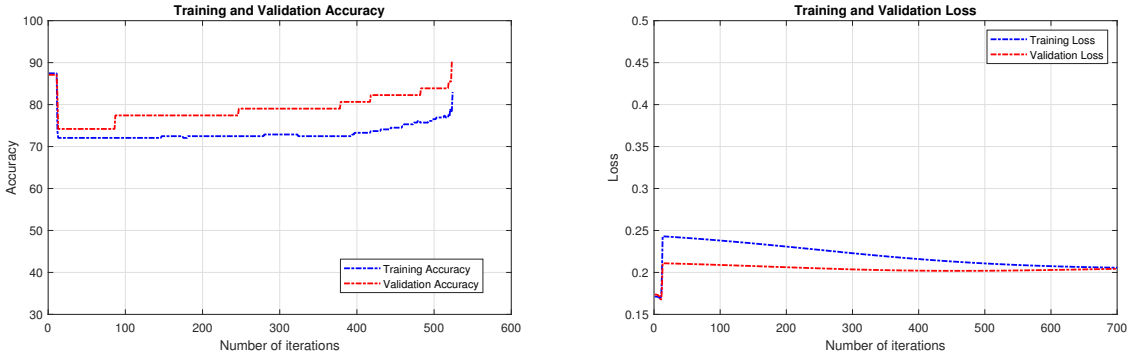

Figure 2: The left panel shows the accuracy graph, and the right panel shows the loss graph of the iterations of Algorithm 1 (RLSP-L_1).

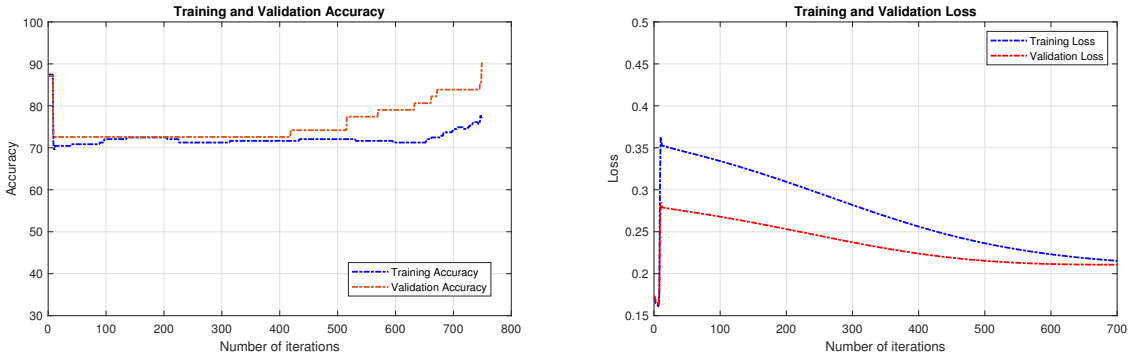

Figure 3: The left panel shows the accuracy graph, and the right panel shows the loss graph of the iterations of Algorithm 1 (RLSP-L_2).

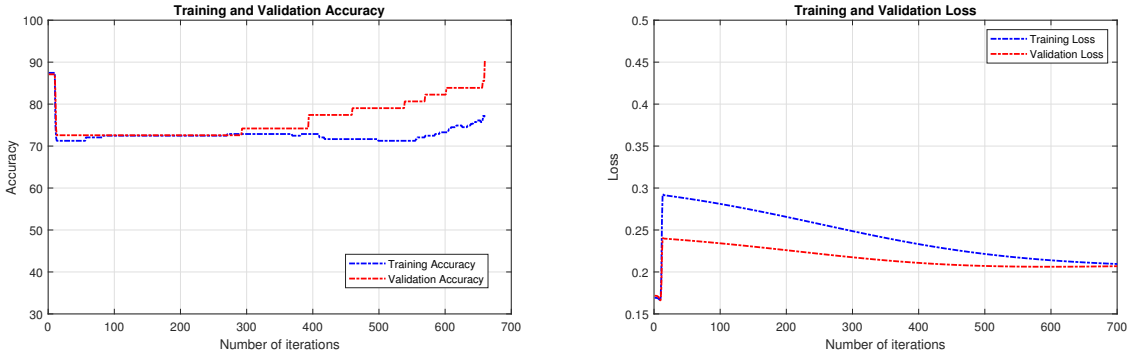

Figure 4: The left panel shows the accuracy graph, and the right panel shows the loss graph of the iterations of Algorithm 1 (RLSP-L_1-CL_1).

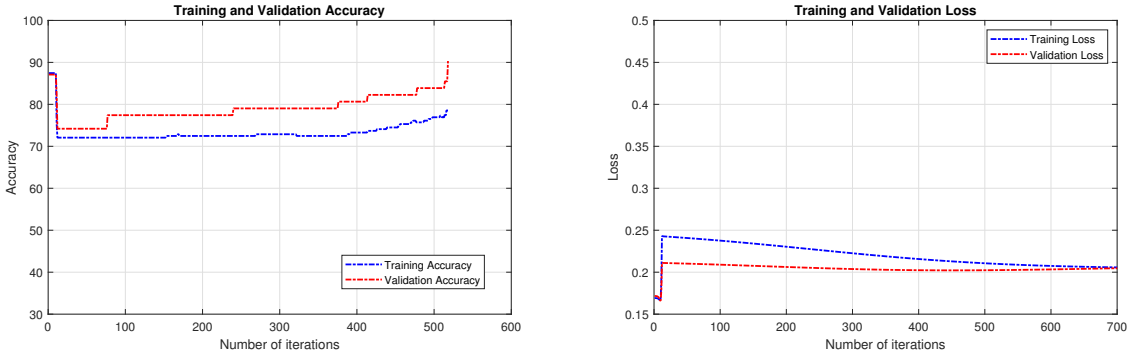

Figure 5: The left panel shows the accuracy graph, and the right panel shows the loss graph of the iterations of Algorithm 1 (RLSP-L_2-CL_2).

Remark 5.1. From Figures 2-5, we observe the following.

(i) The training accuracy steadily increases across iterations, indicating that the models are progressively learning from the training data. Concurrently, the validation accuracy also exhibits a rising trend, reflecting that the models are learning from the training data and generalizing well to unseen validation data.

(ii) This increasing trend suggests that the algorithms effectively fit the training data and enhance their predictive capabilities with more iterations. The proximity between the training and validation accuracy curves indicates minimal overfitting, a desirable characteristic in machine learning models.

(iii) The training loss consistently decreases as the number of iterations increases, a sign that the models are effectively minimizing the error on the training dataset. Equally important, the validation loss follows a decreasing trend, a positive indicator that the models are not overfitting and are performing well on the validation data.

(iv) The plateau observed towards the end of the iterations indicates that the models have reached a point where further training does not significantly reduce the training loss, implying convergence. The validation loss curve closely follows the training loss curve, further confirming the generalization capability of the models.

(v) The minimal gaps between training and validation accuracy and loss curves are a key observation, indicating the models' strong generalization abilities. This suggests that the models are not overfitting and can handle unseen data effectively.

From Remark 5.1, we can conclude that the analysis of Figures 2-5 underscores the effectiveness and reliability of the proposed bi-inertial Mann projective forward-backward splitting algorithms. These algorithms demonstrate superior learning efficiency and generalization performance, making them valuable tools for solving complex variational inclusion problems and practical applications such as lung cancer classification.

6. Conclusions

In this article, we investigated the bi-inertial Mann projective forward-backward splitting algorithm to solve variational inclusion problems within real Hilbert spaces. We established a weak convergence theorem for the sequence of iterates generated by the proposed algorithm under suitable conditions. This theoretical foundation ensures that our algorithm is robust and reliable.

Additionally, we conducted extensive numerical experiments in infinite-dimensional spaces to demonstrate the efficiency and effectiveness of our proposed algorithm compared to existing methods in the literature. The numerical results indicate that our algorithm not only converges rapidly but also requires fewer iterations to reach the solution, thereby outperforming other methods in terms of computational efficiency.

To further validate the practical applicability of our algorithm, we applied it to data classification tasks using a lung cancer dataset. Our algorithm achieved the highest testing accuracy, as detailed in Table 5, surpassing the performance of other algorithms. The accuracy and loss graphs presented in Figures 2-5 show that our algorithm maintains a strong performance across iterations without displaying signs of overfitting. This stability is crucial for ensuring consistent and reliable classification results in real-world applications.

In conclusion, our bi-inertial Mann projective forward-backward splitting algorithm offers a powerful and efficient solution for variational inclusion problems. Its theoretical robustness, computational efficiency, and practical effectiveness in data classification tasks highlight its potential for a wide range of applications in applied mathematics, signal processing, and machine learning. Future research can explore further optimizations and extensions of this algorithm to address even more complex and diverse problems.

Availability of data

The lung cancer dataset used in this work was obtained from publicly available dataset: https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer.

Author contributions

Writing original draft and Software: P. P.; writing review and editing and supervision: W. C.; writing review and software: K. S.. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was supported by the National Research Council of Thailand and University of Phayao (N42A650334), and University of Phayao and Thailand Science Research and Innovation Fund (Fundamental Fund 2025, Grant No. 5013/2567).

References

[1] Y. Alber, I. Ryazantseva, *Nonlinear ill-posed problems of monotone type*, Springer, Dordrecht, (2006). 1
[2] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal., **9** (2001), 3–11.
[3] H. H. Bauschke, P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, (2011). 2.6
[4] L.-C. Ceng, Q. H. Ansari, M. M. Wong, J.-C. Yao, *Mann type hybrid extragradient method for variational inequalities, variational inclusions and fixed point problems*, Fixed Point Theory, **13** (2012), 403–422. 1
[5] S. Chen, D. Heng, J. Huang, Z. Chen, J Zhao, *Two-step inertial adaptive iterative algorithm for solving the split common fixed point problem of directed operators*, J. Nonlinear Funct. Anal., **2023** (2023), 12 pages. 1
[6] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, **53** (2004), 475–504.
[7] P. L. Combettes, V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., **4** (2005), 1168–1200. 1
[8] K. Goebel, W. A. Kirk, *Topics in metric fixed point theory*, Cambridge University Press, Cambridge, (1990).
[9] N. X. Hai, P. Q. Khanh, *The solution existence of general variational inclusion problems*, J. Math. Anal. Appl., **328** (2007), 1268–1277. 1
[10] N.-J. Huang, *A new completely general class of variational inclusions with noncompact valued mappings*, Comput. Math. Appl., **35** (1998), 9–14. 1
[11] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, *Extreme learning machine: Theory and applications*, Neurocomputing, **70** (2006), 489–501. 5
[12] O. S. Iyiola, Y. Shehu, *Convergence results of two-step inertial proximal point algorithm*, Appl. Numer. Math., **182** (2022), 57–75. 1
[13] L. O. Jolaoso, Y. Shehu, J. C. Yao, R. Xu, *Double inertial parameters forward-backward splitting method: Applications to compressed sensing, image processing, and SCAD penalty problems*, J. Nonlinear Var. Anal., **7** (2023), 627–646. 1

[14] W. Khuangsatung, A. Kangtunyakarn, *Algorithm of a new variational inclusion problem and strictly pseudononspreading mapping with application*, Fixed Point Theory Appl., **2014** (2014), 27 pages. 1
[15] P.-L. Lions, B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., **16** (1979), 964–979. 1
[16] G. López, V. Martín-Márquez, F. Wang, H.-K. Xu, *Forward-backward splitting methods for accretive operators in Banach spaces*, Abstr. Appl. Anal., **2012** (2012), 25 pages.
[17] D. A. Lorenz, T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vis., **51** (2015), 311–325. 1
[18] Y. Malitsky, M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., **30** (2020), 1451–1472. 1
[19] G. Marino, H. K. Xu, *Convergence of generalized proximal point algorithms*, Commun. Pure Appl. Anal., **3** (2004), 791–808. 2
[20] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., **29** (1962), 341–346. 2
[21] N. Pakkaranang, *Double inertial extragradient algorithms for solving variational inequality problems with convergence analysis*, Math. Methods Appl. Sci., **47** (2024), 11642–11669. 1
[22] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys., **4** (1964), 1–17. 1
[23] S. Raschka, *Model evaluation, model selection, and algorithm selection in machine learning*, arXiv preprint arXiv:1811.12808, (2018), 1–49. 5
[24] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., **14** (1976), 877–898. 2.1
[25] B. Tan, S. Y. Cho, *Strong convergence of inertial forward–backward methods for solving monotone inclusions*, Appl. Anal., **101** (2022), 5386–5414. 1
[26] B. Tan, X. Qin, *On relaxed inertial projection and contraction algorithms for solving monotone inclusion problems*, Adv. Comput. Math., **50** (2024), 35 pages. 1
[27] B. Tan, X. Qin, J.-C. Yao, *Strong convergence of self-adaptive inertial algorithms for solving split variational inclusion problems with applications*, J. Sci. Comput., **87** (2021), 34 pages. 1
[28] R. U. Verma, *A-monotonicity and applications to nonlinear variational inclusion problems*, J. Appl. Math. Stoch. Anal., **2004** (2004), 193–195. 1
[29] S.-S. Zhang, J. H. W. Lee, C. K. Chan, *Algorithms of common solutions to quasi variational inclusion and fixed point problems*, Appl. Math. Mech. (English Ed.), **29** (2008), 571–581. 1
[30] C. Zhang, Y. Wang, *Proximal algorithm for solving monotone variational inclusion*, Optimization, **67** (2018), 1197–1209. 1