FIXED POINT METHODS AND OPTIMIZATION
Volume 1 (2024), No. 2, 88–100
https://doi.org/10.69829/fpmo-024-0102-ta01
Tulipa Opera Scholarum

A PROJECTIVE DOUBLE INERTIAL ISHIKAWA FORWARD-BACKWARD SPLITTING ALGORITHM FOR VARIATIONAL INCLUSION PROBLEMS WITH APPLICATION TO OSTEOPOROSIS PREDICTION

WATCHARAPORN CHOLAMJIAK[1,*]

[1] *School of Science, University of Phayao, Phayao 56000, Thailand*

Abstract. This work introduces a projective double inertial Ishikawa forward-backward splitting algorithm for solving variational inclusion problems in Hilbert spaces. We establish a weak convergence theorem under suitable control conditions, ensuring the reliability of the proposed approach. Numerical experiments, including an example in an infinite-dimensional space, demonstrate the algorithm's efficiency and validate the theoretical results. Furthermore, our study shows the effectiveness of applying the proposed algorithm to osteoporosis prediction using a multi-layer ELM, with the 2-layer ELM configuration achieving the highest performance across all metrics (accuracy, precision, recall, and F1-score), underscoring its robustness and efficiency.

Keywords. Variational inclusion problem, inertial method, weak convergence, multi-layer ELM, osteoporosis.

1. Introduction and Preliminaries

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ associated with the norm $\| \cdot \|$. Suppose $\mathbb{F} : \mathcal{H} \to \mathcal{H}$ is a single-valued monotone operator, and $\mathbb{G} : \mathcal{H} \to 2^{\mathcal{H}}$ is a set-valued operator. In this work, we address the variational inclusion problem (VIP) in a real Hilbert space, which seeks to find $\hat{f} \in \mathcal{H}$ such that

$$0 \in (\mathbb{F} + \mathbb{G})\hat{f}. \tag{1.1}$$

The variational inclusion problem (VIP) encompasses a wide range of problems, including convex minimization, equilibrium, variational inequality, and split feasibility problems [4, 7, 8, 17]. Additionally, VIP has extensive applications across various fields such as signal processing, image reconstruction, optimal control, quantum mechanics, and machine learning (see [2, 6, 9, 10, 16]).

The resolvent operator $J_\gamma^{\mathbb{G}}$ associated with $\mathbb{G}$ and γ is the mapping $J_\gamma^{\mathbb{G}} : \mathcal{H} \to \mathcal{H}$ defined by

$$J_\gamma^{\mathbb{G}}(f) := (I + \gamma\mathbb{G})^{-1}(f), \ \forall f \in \mathcal{H},$$

where γ is a positive number and I denotes the identity operator on $\mathcal{H}$, it is established that the resolvent operator $J_\gamma^{\mathbb{G}}(f)$ is single-valued, nonexpansive, and 1-inverse strongly monotone [5]. Furthermore, the solution to the problem (1.1) is a fixed point of the operator $J_\gamma^{\mathbb{G}}(I - \gamma\mathbb{F})$ for all $\gamma > 0$ [14].

As a way to speed up the convergence algorithm, Moudafi and Oliny proposed an inertial forward-backward splitting algorithm (IFBSA) for approximating a solution of the VIP (1.1) using an inertial

*Corresponding author.
E-mail address: watcharaporn.ch@up.ac.th (W. Cholamjiak)
2020 Mathematics Subject Classification: 46E20, 52A07, 68Q04.
Accepted November 17, 2024.

technique. For two given points f^0 and f^1, the sequence $\{f^k\}$ is defined as follows:

$$\begin{cases} h^k = f^k + \alpha^k(f^k - f^{k-1}), \\ f^{k+1} = J^{\mathbb{G}}_{\gamma^k}(h^k - \gamma^k \mathbb{F}h^k), \ k \geq 1. \end{cases} \tag{1.2}$$

They obtained a weak convergence result using algorithm (1.2), provided that $\gamma^k < \frac{2}{L}$, with L being the Lipschitz constant of $\mathbb{F}$.

Recently, Jun-On and Cholamjiak [13] introduced a double relaxed inertial projective forward- backward splitting algorithm for variational inclusion problems, incorporating a projection onto a nonempty closed and convex set $\mathcal{P}$. The algorithm is applied to an asymmetrical educational dataset of students from 109 schools, utilizing nine asymmetric attributes as inputs to predict students' mathematical integrated skills. Its performance is compared with other algorithms in the literature to demonstrate its effectiveness. This algorithm was generated as $f^{-1}, f^0, f^1 \in \mathcal{H}$ and

$$\begin{cases} h^k = f^k + \theta^k(f^k - f^{k-1}) + \delta^k(f^{k-1} - f^{k-2}), \\ g^k = (1-\alpha^k)f^k + \alpha^k J^{\mathbb{G}}_{\gamma^k}(I - \gamma^k \mathbb{F})h^k, \\ f^{k+1} = P_{\mathcal{P}}(J^{\mathbb{G}}_{\gamma^k}(I - \gamma^k \mathbb{F})g^k), \ k \geq 1. \end{cases} \tag{1.3}$$

where $\{\alpha^k\} \subset (a,b) \subset (0,1], \{\gamma^k\} \subset (c,d) \subset (0, 2\beta)$, $\{\theta^k\}, \{\delta^k\} \subset (-\infty, \infty)$ such that β is the coefficient of the operator $\mathbb{F}$ and the $\{\theta^k\}$ and δ^k satisfies the following conditions:

$$\sum_{k=1}^{\infty} |\theta^k| \|f^k - f^{k-1}\| < \infty \text{ and } \sum_{k=1}^{\infty} |\delta^k| \|f^{k-1} - f^{k-2}\| < \infty.$$

Building on previous work, we propose an algorithm that integrates a double inertial technique with two iterative steps and a projection method for the forward-backward splitting algorithm to address variational inclusion problems in Hilbert spaces. Under suitable conditions, we establish a weak convergence theorem for the proposed approach. Additionally, we provide an example in an infinite-dimensional space to illustrate the validity of our main theorem. As an application, we apply the algorithm in a machine learning context to predict osteoporosis.

Now, we present definitions and lemmas used in this article.

Definition 1.1. Suppose that $\mathbb{F} : \mathcal{H} \to \mathcal{H}$ is said to be

1. $\mathbb{F}$ is a monotone mapping if the following hold:

$$\langle \mathbb{F}f - \mathbb{F}h, f - h \rangle \geq 0.$$

2. $\mathbb{F}$ is L-Lipschitz continuous if there is a constant $L > 0$, as follows:

$$\|\mathbb{F}f - \mathbb{F}h\| \leq L\|f - h\|.$$

If $L = 1$, then $\mathbb{F}$ is called nonexpansive.

3. $\mathbb{F}$ is firmly nonexpansive if

$$\|\mathbb{F}f - \mathbb{F}h\|^2 \leq \|f - h\|^2 - \|(I - \mathbb{F})f - (I - \mathbb{F})h\|^2,$$

or equivalently

$$\langle \mathbb{F}f - \mathbb{F}h, f - h \rangle \geq \|\mathbb{F}f - \mathbb{F}h\|^2.$$

4. β-cocoercive or β-inverse strongly monotone if $\beta\mathbb{F}$ is firmly nonexpansive when $\beta > 0$.

Lemma 1.2. [11] *Assume that $\mathbb{F} : \mathcal{H} \to \mathcal{H}$ is a nonexpansive mapping with $Fix(\mathbb{F}) \neq \phi$. If there exists a sequence $\{f^k\}$ in $\mathcal{H}$, the following implications hold: $f^k \rightharpoonup f \in \mathcal{H}$ and $\|f^k - \mathbb{F}f^k\| \to 0$ $\Longrightarrow f \in Fix(\mathcal{F})$.*

Lemma 1.3. [15] *Let $\mathbb{F} : \mathcal{H} \to \mathcal{H}$ be β-cocoercive mapping and $\mathbb{G} : \mathcal{H} \to 2^{\mathcal{H}}$ a maximal monotone mapping. Then, we have*

1. for $\gamma > 0$, $Fix(J^{\mathbb{G}}_{\gamma}(I - \gamma\mathbb{F})) = (\mathbb{F} + \mathbb{G})^{-1}(0)$;

2. for $0 < \gamma < \overline{\gamma}$ and $f \in \mathcal{H}$, $\|f - J^{\mathbb{G}}_{\gamma}(I - \gamma\mathbb{F})f\| \leq 2\|f - J^{\mathbb{G}}_{\overline{\gamma}}(I - \overline{\gamma}\mathbb{F})f\|$.

Lemma 1.4. [18] *Let Ψ be a nonempty subset of $\mathcal{H}$ and $\{f^k\}$ be a sequence in $\mathcal{H}$. Then, the following conditions holds:*

1. For all $f \in \Psi$, $\lim_{k\to\infty} \|f^k - f\|$ exists.

2. Every sequential weak cluster point of $\{f^k\}$ belongs to Ψ.

Then $\{f^k\}$ converges weakly to a point in Ψ.

2. Weak Convergence Results

In this section, let $\mathcal{P}$ be a nonempty, closed and convex subset of a Hilbert space $\mathcal{H}$. Let $\mathbb{F} : \mathcal{H} \to \mathcal{H}$ be a β-inverse strongly monotone operator and $\mathbb{G} : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximal monotone operator such that $(\mathbb{F} + \mathbb{G})^{-1}(0) \cap \mathcal{P} \neq \emptyset$, and let $P_{\mathcal{P}}$ be a metric projection on $\mathcal{P}$.

Algorithm 2.1. *Projective Double Inertial Ishikawa Forward-Backward Splitting algorithm (PDI-IFBS)*
Initialization: *Select arbitrary points $f^0, h^{-1}, h^0 \in \mathcal{H}$, $\{\alpha^k\}, \{\beta^k\} \subset (a, b] \subset (0, 1]$, $\{\gamma^k\} \subset (c, d) \subset (0, 2\beta)$, and $\{\theta^k\}, \{\delta^k\} \subset (-\infty, \infty)$.*
Iterative Steps: *Set $k = 0$, compute $\{f^{k+1}\}$ as following steps:*
Step 1: *Compute*

$$g^k = (1 - \alpha^k)f^k + \alpha^k J^k f^k.$$

Step 2: *Compute*

$$h^{k+1} = (1 - \beta^k)f^k + \beta^k J^k g^k.$$

Step 3: *Compute*

$$f^{k+1} = P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})),$$

where $J^k = J^{\mathbb{G}}_{\gamma^k}(I - \gamma^k\mathbb{F})$. Set $k = k + 1$ and return to **Step 1**.

Theorem 2.2. *The sequence $\{f^k\}$ generated by Algorithm 2.1 converges weakly to an element in $(\mathbb{F} + \mathbb{G})^{-1}(0) \cap \mathcal{P}$. Suppose also the conditions below hold:*

$$\sum_{k=1}^{\infty} |\theta^k| \|h^{k+1} - h^k\| < \infty \text{ and } \sum_{k=1}^{\infty} |\delta^k| \|h^k - h^{k-1}\| < \infty.$$

Proof. Let $f^* \in (\mathbb{F} + \mathbb{G})^{-1}(0) \cap \mathcal{P}$. Since $\{\gamma^k\} \subset (0, 2\beta)$, J^k is nonexpansive mapping by [21] and science $P_{\mathcal{P}}$ is also nonexpansive, we have

$$\begin{aligned}
\|f^{k+1} - f^*\| &= \|P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})) - f^*\| \\
&\leq \|h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}) - f^*\| \\
&\leq \|h^{k+1} - f^*\| + |\theta^k|\|h^{k+1} - h^k\| + |\delta^k|\|h^k - h^{k-1}\| \\
&= \|(1 - \beta^k)f^k + \beta^k J^k g^k - f^*\| + |\theta^k|\|h^{k+1} - h^k\| + |\delta^k|\|h^k - h^{k-1}\| \\
&\leq (1 - \beta^k)\|f^k - f^*\| + \beta^k\|J^k g^k - f^*\| + |\theta^k|\|h^{k+1} - h^k\| + |\delta^k|\|h^k - h^{k-1}\| \\
&\leq (1 - \beta^k)\|f^k - f^*\| + \beta^k\|g^k - f^*\| + |\theta^k|\|h^{k+1} - h^k\| + |\delta^k|\|h^k - h^{k-1}\| \\
&= (1 - \beta^k)\|f^k - f^*\| + \beta^k\|(1 - \alpha^k)f^k + \alpha^k J^k f^k - f^*\| + |\theta^k|\|h^{k+1} - h^k\| \\
&\quad + |\delta^k|\|h^k - h^{k-1}\| \\
&\leq (1 - \beta^k)\|f^k - f^*\| + \beta^k\big((1 - \alpha^k)\|f^k - f^*\| + \alpha^k\|J^k f^k - f^*\|\big) + |\theta^k|\|h^{k+1} \\
&\quad - h^k\| + |\delta^k|\|h^k - h^{k-1}\|
\end{aligned}$$

$$\leq \|f^k - f^*\| + |\theta^k|\|h^{k+1} - h^k\| + |\delta^k|\|h^k - h^{k-1}\|$$

By our conditions, it follows form the Lemma in [3], we obtain $\lim_{k\to\infty} \|f^k - f^*\|$ exists. This implies that $\{f^k\}$ is bounded. So, the sequences $\{g^k\}$ and $\{h^k\}$ are also bounded. Since $J^{\mathbb{G}}_{\tau^k}$ is a firmly nonexpansive mapping, let $\mathcal{K}^k = h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})$ then we have

$$\begin{aligned}
& \|f^{k+1} - f^*\|^2 \\
= \;& \|P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})) - f^*\|^2 \\
\leq \;& \|h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}) - f^*\|^2 \\
\leq \;& \|h^{k+1} - f^*\|^2 + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
= \;& \|(1-\beta^k)f^k + \beta^k J^k g^k - f^*\|^2 + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\|J^k g^k - f^*\|^2 + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\|g^k - f^*\|^2 + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
= \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\|(1-\alpha^k)f^k + \alpha^k J^k f^k - f^*\|^2 \\
& + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\big((1-\alpha^k)\|f^k - f^*\|^2 + \alpha^k\|J^k f^k - f^*\|^2\big) \\
& + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\Big((1-\alpha^k)\|f^k - f^*\|^2 + \alpha^k\big(\|f^k - \gamma^k\mathbb{F}f^k - f^* + \gamma^k\mathbb{F}f^*\|^2 \\
& - \|f^k - \gamma^k\mathbb{F}f^k - J^k f^k - f^* + \gamma^k\mathbb{F}f^* + J^k f^*\|^2\big)\Big) \\
& + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& (1-\beta^k)\|f^k - f^*\|^2 + \beta^k\Big((1-\alpha^k)\|f^k - f^*\|^2 + \alpha^k\big(\|f^k - f^*\|^2 - 2\gamma^k\langle f^k - f^*, \mathbb{F}f^k - \mathbb{F}f^*\rangle \\
& + (\gamma^k)^2\|\mathbb{F}f^k - \mathbb{F}f^*\|^2 - \|f^k - \gamma^k\mathbb{F}f^k - J^k f^k - f^* + \gamma^k\mathbb{F}f^* + J^k f^*\|^2\big)\Big) \\
& + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
\leq \;& \|f^k - f^*\|^2 + 2\langle \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1}), \mathcal{K}^k - f^*\rangle \\
& - \beta^k\alpha^k\Big(\gamma^k(2\beta - \gamma^k)\|\mathbb{F}f^k - \mathbb{F}f^*\|^2 + \|f^k - \gamma^k\mathbb{F}f^k - J^k f^k - f^* + \gamma^k\mathbb{F}f^* + J^k f^*\|^2\Big).
\end{aligned}$$

This implies that

$$\begin{aligned}
& \beta^k\alpha^k\Big(\gamma^k(2\beta - \gamma^k)\|\mathbb{F}f^k - \mathbb{F}f^*\|^2 + \|f^k - \gamma^k\mathbb{F}f^k - J^k f^k + \gamma^k\mathbb{F}f^*\|^2\Big) \qquad (2.1) \\
\leq \;& \|f^k - f^*\|^2 - \|f^{k+1} - f^*\|^2 + 2\langle \theta^k(b^{k+1} - b^k) + \delta^k(b^k - b^{k-1}), c^{k+1} - a^*\rangle.
\end{aligned}$$

It follows from our conditions and $\lim_{k\to\infty} \|f^k - f^*\|$ exists, we have

$$\lim_{k\to\infty} \|\mathbb{F}f^k - \mathbb{F}f^*\| = \lim_{k\to\infty} \|f^k - \gamma^k\mathbb{F}f^k - J^k f^k + \gamma^k\mathbb{F}f^*\| = 0.$$

This implies that

$$\lim_{k\to\infty} \|f^k - J^k f^k\| = 0. \qquad (2.2)$$

Since $\liminf_{k\to\infty} \gamma^k > 0$, there is $\gamma > 0$ such that $\gamma^k > \gamma$. Lemma 1.3 (ii), we obtain

$$\|f^k - J^{\mathbb{G}}_{\gamma}(I - \gamma\mathbb{F})f^k\| \leq 2\|f^k - J^k f^k\| = 0. \tag{2.3}$$

Then, by (2.2) and (2.3), we obtain

$$\lim_{k\to\infty} \|f^k - J^{\mathbb{G}}_{\gamma}(I - \gamma\mathbb{F})f^k\| = 0.$$

Since $\{f^k\}$ is bound, we can let $\tilde{f}$ be a sequential weak cluster point of $\{f^k\}$. By Lemma 1.2, we deduce that $\tilde{f}$ belongs to the fixed point set $\mathrm{Fix}(J^{\mathbb{G}}_{\gamma}(I-\gamma\mathbb{F}))$, which is equivalent to the set $(\mathbb{F}+\mathbb{G})^{-1}(0)$. Given that $\tilde{f}$ is a sequence in $\mathcal{P}$, and $\mathcal{P}$ is a nonempty, closed, and convex subset, it follows that $\tilde{f} \in (\mathbb{F}+\mathbb{G})^{-1}(0)\cap\mathcal{P}$. By employing Opial's Lemma (Lemma 1.3), we can establish that $\tilde{f}$ weakly converges to an element in $(\mathbb{F}+\mathbb{G})^{-1}(0) \cap \mathcal{P}$. □

Remark 2.3. (i) If $\alpha^k = 1$, Algorithm 1 reduces to following algorithm:
Projective Double Inertial Two-step Forward-Backward Splitting algorithm I (PDI-TFBS(I)):

$$\begin{cases} g^k = J^k f^k, \\ h^{k+1} = (1-\beta^k)f^k + \beta^k J^k g^k, \\ f^{k+1} = P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})); \end{cases}$$

(ii) If $\beta^k = 1$, Algorithm 1 reduces to following algorithm:
Projective Double Inertial Two-step Forward-Backward Splitting algorithm II (PDI-TFBS(II)):

$$\begin{cases} g^k = (1-\alpha^k)f^k + \alpha^k J^k f^k, \\ h^{k+1} = J^k g^k, \\ f^{k+1} = P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})); \end{cases}$$

(iii) If $\alpha^k = 1$ and $\beta^k = 1$, Algorithm 1 reduces to following algorithm:
Projective Double Inertial Two-step Forward-Backward Splitting algorithm III (PDI-TFBS(III)):

$$\begin{cases} g^k = J^k f^k, \\ h^{k+1} = J^k g^k, \\ f^{k+1} = P_{\mathcal{P}}(h^{k+1} + \theta^k(h^{k+1} - h^k) + \delta^k(h^k - h^{k-1})); \end{cases}$$

We shall give an example in the infinitely dimension space $L_2[0,1] = \{f(t) : \int_0^1 f(t)dt < \infty\}$ where such that $\|.\|$ is L_2-norm defined by $\|f\| = \sqrt{\int_0^1 |a(t)|^2 dt}$ and the inner product $\langle f, h\rangle = \int_0^1 f(t)h(t)dt$ for supporting our main theorem.

Example 2.4. Let $\mathcal{H} = L_2[0,1]$, $\mathbb{F}f(t) = 3f(t)$ and $\mathbb{G}f(t) = 4f(t)$ where $f(t) \in L_2[0,1]$.

In our experiments, we use the Cauchy errors $\|f^k - f^{k+1}\|^2 < 10^{-5}$ to stop the iteration. To meet the highest performance of the our algorithm, we shall consider the necessary parameters of the algorithm when

$$\theta^k = \begin{cases} \frac{0.99}{\|h^{k+1}-h^k\|^3} & \text{, if } h^{k+1} \neq^k \text{ and } k > M \\ 0.99 & \text{, otherwise} \end{cases} \tag{2.4}$$

and

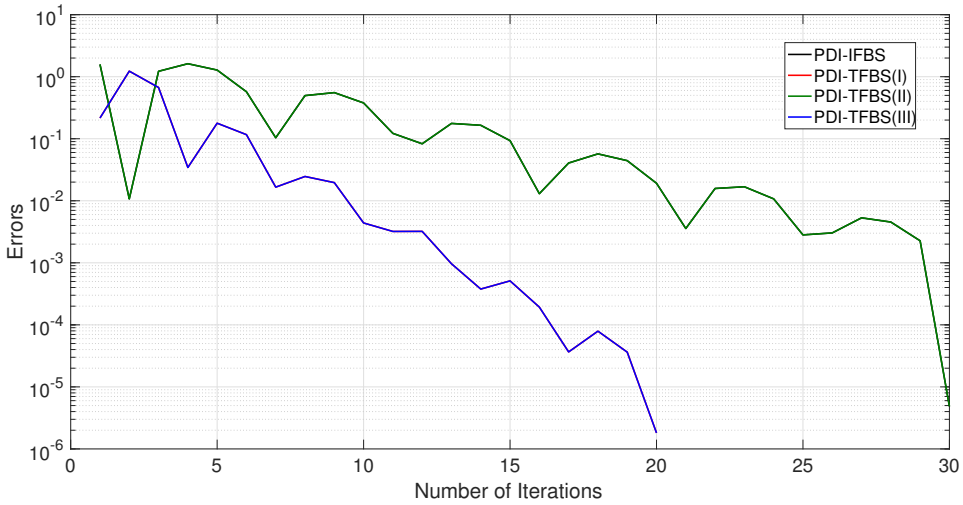
$$\delta^k = \begin{cases} \frac{-0.1}{\|h^k-h^{k-1}\|^3} & \text{, if } h^k \neq h^{k-1} \text{ and } k > M \\ -0.1 & \text{, otherwise} \end{cases} \tag{2.5}$$

where M is the iteration number that we want to stop, and we choose $\gamma^k = \frac{1.99}{3}$, $\alpha^k = \beta^k = \frac{k}{1.5k+1}$. Table 1 shows the numerical results for difference parameters of our algorithm and Cauchy errors of

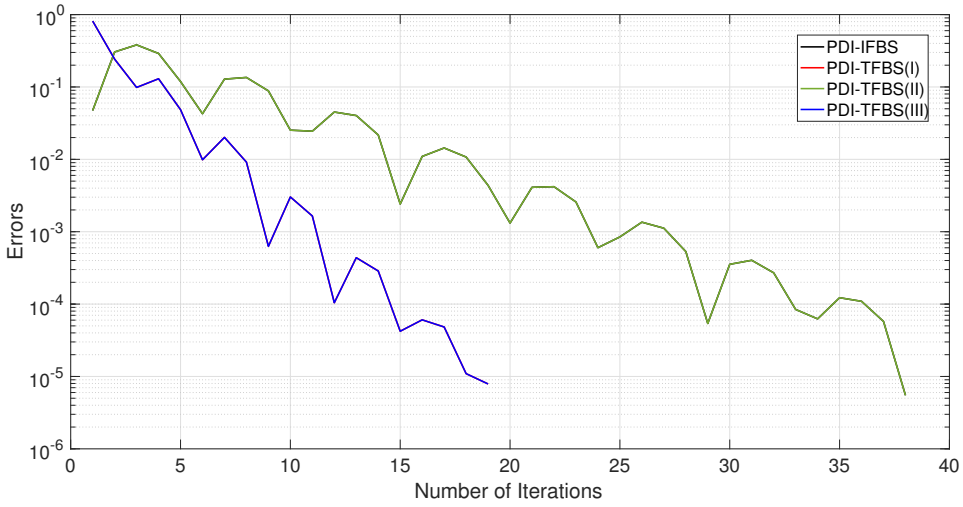
them are show in Figure 1.

Table 1. Numerical results for difference parameters of our algorithm.

Initialization	Algorithm	Number of Iterations	CPU Time
$f^0 = t^2 - 4t$, $h^{-1} = \frac{\sin(t)}{2}$,	PDI-IFBS	30	3.9676
	PDI-TFBS(I)	20	2.6233
$h^0 = \sin(t)$, $\mathcal{P} = \{f : \langle 2t^2, f\rangle \leq 5\}$	PDI-TFBS(II)	30	3.8276
	PDI-TFBS(III)	20	2.6901
$f^0 = t + \log(t+1)^2$, $h^{-1} = \sin(t+1)$,	PDI-IFBS	38	4.9603
	PDI-TFBS(I)	19	2.6120
$h^0 = \sin(t)$, $\mathcal{P} = \{f : \langle 2e^t, f\rangle \leq 5\}$	PDI-TFBS(II)	38	4.8947
	PDI-TFBS(III)	19	2.5999



Figures 1. Cauchy errors plots of all our algorithms in Table 1 for initialization $f^0 = t^2 - 4t$, $h^{-1} = \frac{\sin(t)}{2}$, $h^0 = \sin(t)$ and $\mathcal{P} = \{f : \langle 2t^2, f\rangle \leq 5\}$

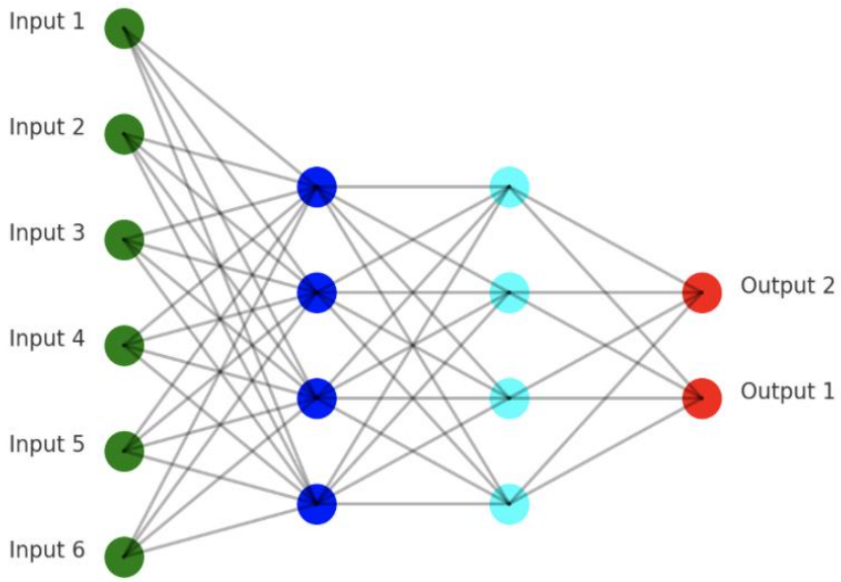


Figures 2. Cauchy errors plots of all our algorithms in Table 1 for initialization $f^0 = t + \log(t+1)^2$, $h^{-1} = \sin(t+1)$, $h^0 = \sin(t)$ and $\mathcal{P} = \{f : \langle 2e^t, f\rangle \leq 5\}$.

From Table 1 and Figures 1-2, errors decrease with increasing iterations across all algorithms, showing convergence. The PDI-TFBS(III) algorithm demonstrates superior error reduction and efficiency performance, particularly for challenging initialization setups. This suggests that the inertial and strategies employed in PDI-TFBS(III) contribute to improved convergence in solving the given problem.

3. Applications

In this section, inspired by the Extreme Learning Machine (ELM) framework initially introduced by Huang et al. [12], we extend the architecture to a 2-layer ELM with two hidden layers, as illustrated in Figure 3. Our newly proposed algorithm serves as the optimizer for this multi-layer ELM, applied to assess osteoporosis risk using a comprehensive dataset from Kaggle (https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factors-influencing-osteoporosis). This dataset provides extensive information on osteoporosis-related health factors, including demographic details, lifestyle choices, medical history, and bone health indicators. The rich dataset supports the development of machine learning models capable of accurately identifying high-risk individuals by analyzing age, gender, hormonal changes, and lifestyle habits. Our approach advances osteoporosis management and prevention strategies by enabling early diagnosis and timely interventions, ultimately reducing fracture risk, improving patient outcomes, and optimizing healthcare resource allocation.



Figures 3. Structure of a feedforward neural network with an input layer consisting of 6 nodes (green), two hidden layers, and an output layer with 2 nodes (red). The first hidden layer contains 4 nodes (dark blue), and the second hidden layer contains 4 nodes (light blue).

In our experiment, the neural network architecture consists of an input layer with 14 nodes (representing the features of the dataset), a first hidden layer with 500 nodes, a second hidden layer with 10 nodes, and an output layer with 2 nodes (representing the target values of the dataset). The dataset is split into 80% for training and 20% for validation.

For the 2-layer Extreme Learning Machine (ELM) process, we consider N distinct samples, where the training set $\mathcal{S} := \{(f^k, t^k) : f^k \in \mathbb{R}^n, t^k \in \mathbb{R}^m, k = 1, 2, \ldots, N\}$ consists of input data f^k and corresponding target outputs t^k. The output function for the first hidden layer (depicted as the blue layer in Figure 3) at the j-th hidden node of the 2-layer ELM with M hidden nodes is mathematically represented as:

$$O_{1j} = \sum_{i=1}^{M} \beta_{1i} \frac{1}{1 + e^{-(w_{1i} f^j + b_{1i})}},$$

where $w_1 i$ is a randomly initialized weight, and $b_1 i$ is a randomly initialized bias for the i-th hidden node. When the hidden matrix of the first layer can be formulated as:

$$H_1 = \begin{bmatrix} \frac{1}{1+e^{-(w_{11} f^1 + b^{11})}} & \cdots & \frac{1}{1+e^{-(w_{1M} f^1 + b_{1M})}} \\ \vdots & \ddots & \vdots \\ \frac{1}{1+e^{-(w_{11} f^N + b_{11})}} & \cdots & \frac{1}{1+e^{-(w_{1M} f^N + b_{1M})}} \end{bmatrix}.$$

Then the output of the first layer will be the input data of the second layer such that

$$H_1 \beta_1 = O_1, \tag{3.1}$$

where $\beta_1 = [\beta_{11}^T, \ldots, \beta_{1M}^T]^T$ is output weight and $O_1 = [O_{11}^T, \ldots, O_{1N}^T]^T$ is the output of the first layer. The output function for the second hidden layer (light blue layer in Figure 3) at k^{th} hidden node of 2-layer ELM with D hidden nodes is mathematically represented as:

$$O_{2k} = \sum_{i=1}^{D} \beta_{2i} \frac{1}{1 + e^{-(w_{2i} O_{1k} + b_{2i})}},$$

where w_{2i} is a randomly initialized weight, and b_{2i} is a randomly initialized bias for the i-th hidden node. The goal is to find the optimal output weights β_{2i} such that

$$H_2 \beta_2 = T, \tag{3.2}$$

where $\beta_2 = [\beta_{21}^T, \ldots, \beta_{2D}^T]^T$ is the vector of optimal output weights of the second layer and $T = [t^{1T}, \ldots, t^{NT}]^T$. When the hidden matrix of the second layer can be formulated as:

$$H_2 = \begin{bmatrix} \frac{1}{1+e^{-(w_{21} O_{11} + b^{21})}} & \cdots & \frac{1}{1+e^{-(w_{2D} O_{11} + b_{2D})}} \\ \vdots & \ddots & \vdots \\ \frac{1}{1+e^{-(w_{21} O_{1N} + b_{21})}} & \cdots & \frac{1}{1+e^{-(w_{2D} O_{1N} + b_{2D})}} \end{bmatrix}.$$

The system of linear equations (3.1) and (3.2) can be solved using a least squares approach, particularly when the Moore-Penrose generalized inverses of H_1 and H_2 are challenging to compute. To prevent overfitting, stabilize the solution, reduce variance, and control model complexity-resulting in models that perform well on new data-we employ the well-known Least Absolute Shrinkage and Selection Operator (LASSO) method [19]: for $\lambda_1, \lambda_2 > 0$

$$\min_{\beta \in R^M} \frac{1}{2} \|H_1 \beta_1 - O_1\|_2^2 + \lambda_1 \|\beta_1\|_1 \tag{3.3}$$

and

$$\min_{\beta_2 \in R^D} \frac{1}{2} \|H_2 \beta_2 - T\|_2^2 + \lambda_2 \|\beta_2\|_1. \tag{3.4}$$

By applying Algorithm 2.1 to solve the problem (3.3) and (3.4), we set $\mathbb{F}\beta \equiv \nabla(\frac{1}{2}\|H_1\beta - O_1\|_2^2)$, $\mathbb{G}\beta \equiv \partial(\lambda_1\|\beta\|_1)$ with $\lambda_1 = 0.01$ for (3.3) and $\mathbb{F}\beta \equiv \nabla(\frac{1}{2}\|H_2\beta - T\|_2^2)$, $\mathbb{G}\beta \equiv \partial(\lambda_2\|\beta\|_1)$ with $\lambda_2 = 0.01$ for (3.4).

We assessed the performance of the classification algorithms using four evaluation metrics: accuracy, precision, recall, and F1-score [20]. These metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%;$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%;$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%;$$

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}.$$

In these formulas, TP represents True Positives, TN True Negatives, FP False Positives, and FN False Negatives.

Additionally, we used binary cross-entropy loss [1] to evaluate the model's ability to distinguish between two classes in binary classification tasks. This loss is computed as the average:

$$\text{Loss} = -\sum_{i=1}^{N} \varphi_i \log \bar{\varphi}_i + (1 - \varphi_i) \log(1 - \bar{\varphi}_i),$$

where $\bar{\varphi}_i$ represents the predicted probability for the i-th instance, φ_i is the corresponding true label, and N is the total number of instances. For the comparison of our Algorithm (2.1) with Algorithm (1.3), we choose $\gamma^k = \frac{0.999}{\|H_1\|^2} / \frac{0.999}{\|H_2\|^2}$, $\alpha^k = 0.8$, $\beta^k = 0.9$, and $\mathcal{P} = \{f : \|f\|^2 \leq 5\}$ where

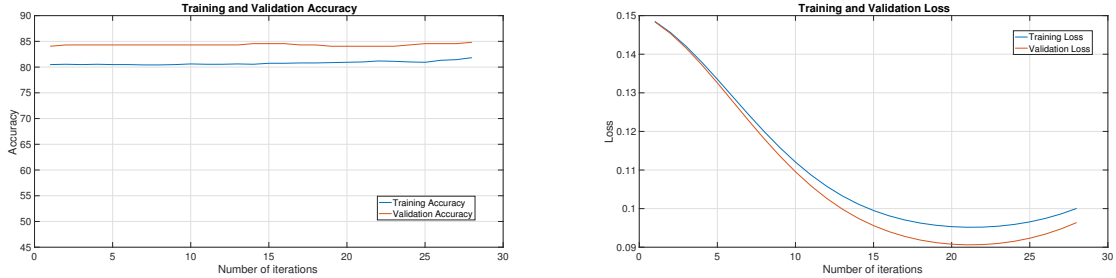
$$\theta^k = \begin{cases} \frac{0.999}{k^2 \|h^{k+1} - h^k\|} & if\ h^{k+1} \neq h^k\ and\ k > N, \\ 0.999 & otherwise; \end{cases}$$

$$\delta^k = \begin{cases} \frac{-0.001}{k^2 \|h^k - h^{k-1}\|} & if\ h^k \neq h^{k-1}\ and\ k > N, \\ -0.001 & otherwise \end{cases}$$

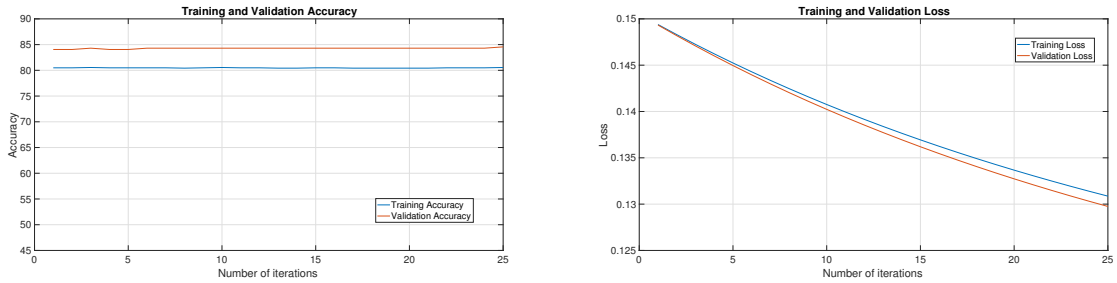
such that N is the number if we want to stop. The results are presented in Table 2.5

Table 2. Numerical results for different configurations of the proposed algorithm applied to 1-Layer and 2-Layer Extreme Learning Machine (ELM) models.

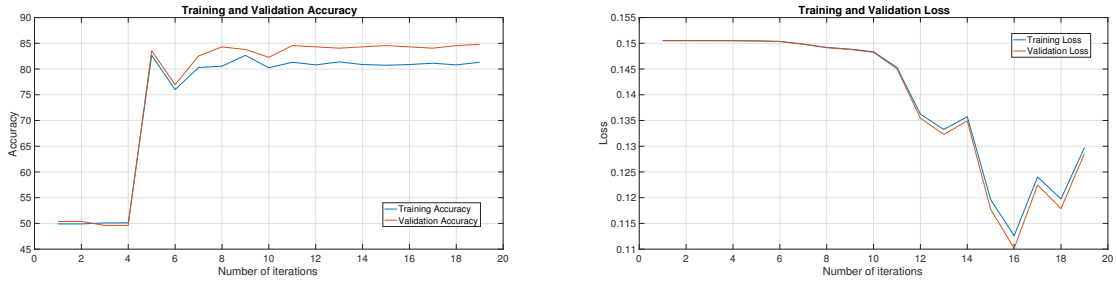
	1-Layer ELM 500 nodes		2-Layer ELM 500 nodes, 10 nodes	
	Algorithm 2.1	Algorithm 1.3	Algorithm 2.1	Algorithm 1.3
Number of Iterations	28	25	19	21
CPU Time	0.9740	0.9072	0.6723	0.7327
Accuracy	84.81	84.56	84.81	84.81
Precision	87.24	85.20	87.24	87.75
Recall	83.00	83.91	83.00	82.69
F1-score	85.07	84.55	85.07	85.14



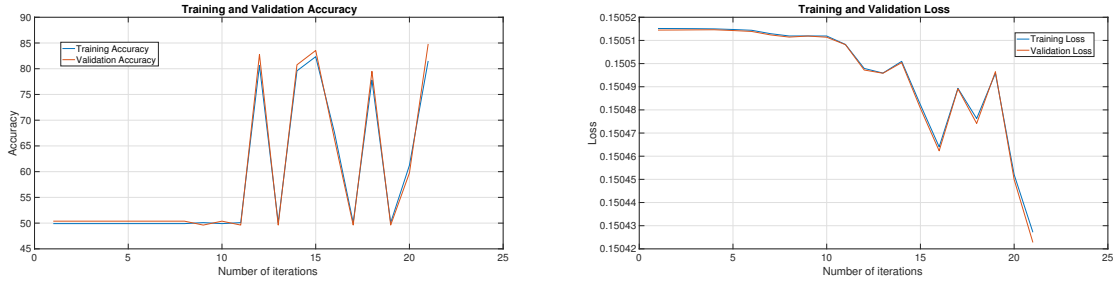
Figures 4-5: Training and validation performance of our Algorithm 2.1 on the 1-layer ELM: (Left) Accuracy stabilizes over iterations, (Right) Loss decreases, showing effective learning.



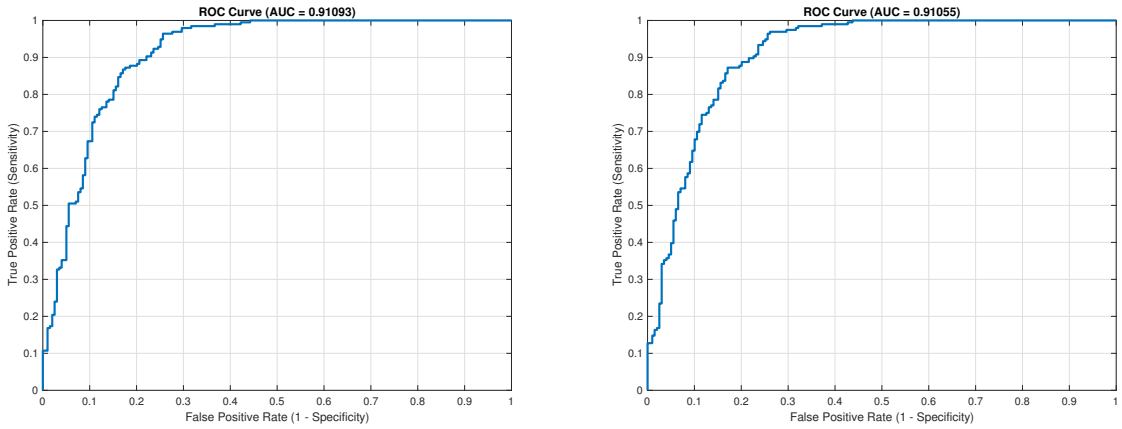
Figures 6-7: Training and validation performance of Algorithm 1.3 on the 1-layer ELM: (Left) Accuracy stabilizes across iterations, (Right) Loss decreases, indicating effective learning and convergence.



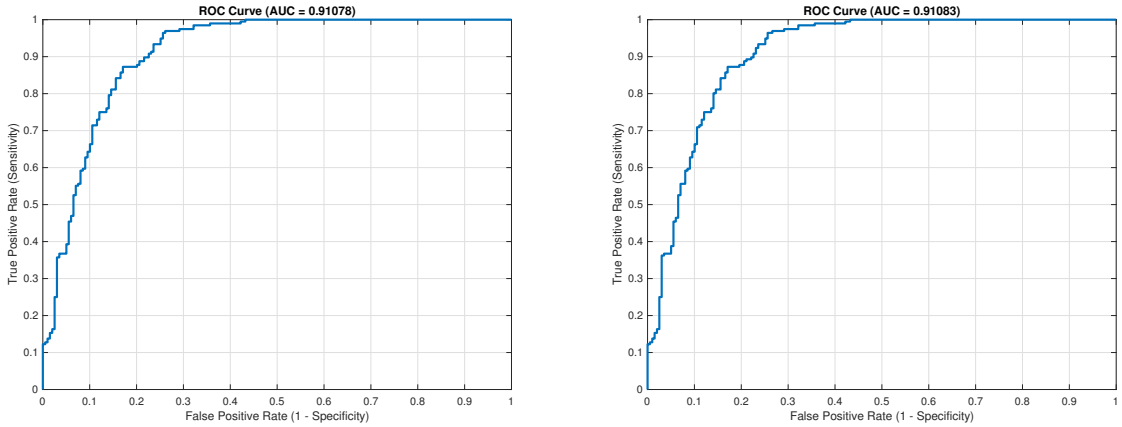
Figures 8-9: Training and validation performance of our Algorithm 2.1 on the 2-layer ELM: (Left) Accuracy rises quickly, then stabilizes; (Right) Loss decreases overall, with fluctuations after 15 iterations, indicating model adaptation.



Figures 10-11: Training and validation performance of Algorithm 1.3 on the 2-layer ELM: (Left) Accuracy shows high fluctuation, (Right) Loss decreases gradually with post-10 iteration fluctuations.



Figures 12-13: ROC curves for our Algorithm 2.1 applied to the 1-layer ELM (left) and 2-layer ELM (right).



Figures 14-15: ROC curves for Algorithm 1.3 applied to the 1-layer ELM (left) and 2-layer ELM (right).

Remark 3.1. (1) From Table 2, the 2-layer ELM models offer improved efficiency (lower iterations and CPU time) while maintaining comparable performance metrics (accuracy, precision, recall, and F1-score) relative to the 1-layer ELM models.
(2) From Figures 4-7, both algorithms demonstrate stable accuracy and decreasing loss, indicating efficient learning and convergence in the 1-layer ELM model.
(3) From Figures 8-11, our Algorithm 2.1 demonstrates stable accuracy and effective learning, while Algorithm 1.3 shows fluctuating accuracy with a slower convergence in the 2-layer ELM model.
(4) From Figures 12-15, both algorithms (2.1 and 1.3) achieve high AUC values in both 1-layer and 2-layer ELM models, reflecting effective classification capabilities.

4. Conclusions

This work introduced a projective double inertial Ishikawa forward-backward splitting algorithm for addressing variational inclusion problems in Hilbert spaces. We proved a weak convergence theorem under suitable control conditions, demonstrating the reliability of our approach. Numerical experiments, including an example in infinite-dimensional space, confirmed the algorithm's efficiency and supported the theoretical findings.

Our study highlights the successful application of this algorithm to osteoporosis prediction using a multi-layer Extreme Learning Machine (ELM). The results indicate that the 2-layer ELM configuration outperformed the 1-layer model across all metrics-accuracy, precision, recall, and F1-score-demonstrating its robustness and efficiency. Specifically, the results from Figures 4-7 show that both algorithms exhibit stable accuracy and a downward trend in loss, suggesting effective learning and convergence in the 1-layer ELM model. In the 2-layer ELM model, Algorithm 2.1 achieved stable accuracy with faster convergence, while Algorithm 1.3 showed more fluctuation and slower convergence, potentially indicating a higher sensitivity to overfitting.

Moreover, both algorithms attained high AUC values in the 1-layer and 2-layer ELM models (Figures 12-15), underscoring their strong classification performance. This study demonstrates that the proposed algorithm is theoretically sound and practical for machine learning applications like osteoporosis prediction, providing a reliable tool for high-dimensional data focusing on minimizing overfitting and maximizing predictive accuracy.

Data Availability

The data are available in Kaggle website (https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factors-influencing-osteoporosis).

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication.

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] M. Akil, R. Saouli, and R. Kachouri. Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Medical Image Analysis*, 63:101692, 2020.

[2] M. Al-Qurashi, S. Rashid, F. Jarad, E. Ali, and R. H. Egami. Dynamic prediction modelling and equilibrium stability of a fractional discrete biophysical neuron model. *Results in Physics*, 48:106405, 2023.

[3] F. Alvarez. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. *SIAM Journal on Optimization*, 14(3):773-782, 2004.

[4] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367-426, 1996.

[5] H. Brézis. Opérateurs maximaux monotones, mathematics studies 5. *Notas de Matematica*, 50, 1973.

[6] C. Byrne. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems*, 18(2):441, 2002.

[7] Y. Censor and T. Elfving. A multiprojection algorithm using Bregman projections in a product space. *Numerical Algorithms*, 8:221-239, 1994.

[8] Y. Censor, A. Gibali, and S. Reich. Algorithms for the split variational inequality problem. *Numerical Algorithms*, 59:301-323, 2012.

[9] P. Chen, J. Huang, and X. Zhang. A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29(2):025011, 2013.

[10] P. L. Combettes. The convex feasibility problem in image recovery. *In Advances in Imaging and Electron Physics*, 95:155-270, 1996.

[11] K. Goebel and W. A. Kirk. *Topics in Metric Fixed Point Theory*, Series No. 28, Cambridge University Press, England, 1990.

[12] G. B. Huang, Q. Y. Zhu, and C. K. Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489-501, 2006.

[13] N. Jun-On and W. Cholamjiak. Enhanced Double Inertial Forward–Backward Splitting Algorithm for Variational Inclusion Problems: Applications in Mathematical Integrated Skill Prediction. *Symmetry*, 16(8):1091, 2024.

[14] B. Lemaire. Which fixed point does the iteration method select?. In: P. Gritzmann, R. Horst, E. Sachs, R. Tichatschke, editors, Recent Advances in Optimization: Proceedings of the 8th French-German Conference on Optimization Trier, pages 154-167, July 21–26, 1996, *Springer, Berlin, Heidelberg*, 1997.

[15] G. López, V. Martín-Márquez, F. Wang, and H. K. Xu. Forward-backward splitting methods for accretive operators in Banach spaces. *In Abstract and Applied Analysis*, Volume 2012, Article ID 109236, Hindawi Publishing Corporation, 2012.

[16] Y. Malitsky and M. K. Tam. A forward-backward splitting method for monotone inclusions without cocoercivity. *SIAM Journal on Optimization*, 30(2):1451-1472, 2020.

[17] A. Moudafi and B. S. Thakur. Solving proximal split feasibility problems without prior knowledge of operator norms. *Optimization Letters*, 8(7):2099-2110, 2014.

[18] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bulletin of the American Mathematical Society, 73:591-597, 1967.

[19] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267-288, 1996.

[20] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo. Cross-validation metrics for evaluating classification performance on imbalanced data. *In 2019 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, pages 14-18. IEEE, Tangerang, Indonesia, 2019.

[21] S. S. Zhang, J. H. Lee, and C. K. Chan.Algorithms of common solutions to quasi variational inclusion and fixed point problems. *Applied Mathematics and Mechanics*, 29(5):571-581, 2008.