CARPATHIAN J. MATH.
Volume **43** (2027), No. 1,
Pages 209-231

Online version at https://www.carpathian.cunbm.utcluj.ro/
Print Edition: ISSN 1584 - 2851; Online Edition: ISSN 1843 - 4401
DOI: https://doi.org/10.37193/CJM.2027.01.13

A Novel Forward-Backward Optimization Scheme for Dementia Screening via Extreme Learning Machines

KUNRADA KANKAM[1], SUPARAT KESORNPROM[2,3], BING TAN[4], WATCHARAPORN CHOLAMJIAK[5,6], PASA SUKSON[6,7], AND PRASIT CHOLAMJIAK[6,7,*]

ABSTRACT. In this paper, we propose a novel self-adaptive forward-backward algorithm for solving convex minimization problems in Hilbert spaces and study its convergence behavior. The algorithm is developed within the forward-backward splitting framework, where a self-adaptive mechanism is introduced to update the control parameters at each iteration. This leads to improve stability and convergence performance. Under appropriate conditions, we establish both weak and strong convergence of the generated sequence. To support the theoretical analysis, we provide numerical experiments in infinite-dimensional Hilbert spaces, demonstrating the efficiency and convergence behavior of the proposed algorithm. Furthermore, to illustrate its practical applicability, it is applied to a classification problem for dementia detection in elderly populations. The algorithm is combined with the Extreme Learning Machine (ELM) in data science. Experimental results, evaluated in terms of accuracy, precision, recall, and F1-score, show that our method achieves superior performance compared to existing approaches.

1. INTRODUCTION AND PRELIMINARIES

Convex minimization problems (CMP) arise naturally in numerous applications across applied mathematics, data science, and engineering, where their rich mathematical structure allows for the development of efficient computational methods. Such problems are widely used in image restoration, signal processing, and machine learning, particularly in data classification tasks involving high-dimensional and structured datasets. In recent years, CMPs have also gained increasing attention in medical applications, where optimization-based models are employed for analyzing patient data, supporting diagnostic classification, and improving decision-making systems. In many of these applications, the problem can be expressed as a composite optimization model involving the sum of two convex functions, where one function is smooth and the other is possibly non-smooth but proximable. This structure has motivated the development of forward-backward type splitting methods and their variants, which have proven to be effective for large-scale and complex problems (see, for instance, [2, 3, 4, 7, 9, 12, 13, 15, 18, 19]).

In this work, we consider the following convex minimization problem. Let $\mathbb{H}$ be a real Hilbert space equipped with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$. Let $F, G : \mathbb{H} \to \mathbb{R} \cup +\infty$ be proper, lower-semicontinuous and convex. We aim to solve the CMP:

$$\min_{e\in\mathbb{H}} \; F(e) + G(e). \tag{1.1}$$

where F is differentiable on $\mathbb{H}$.

The convex minimization problem can be equivalently reformulated as a fixed point problem, which allows the use of operator-theoretic tools for convergence analysis.

Received: 18.06.2026. In revised form: 16.09.2026. Accepted: 17.10.2026
2020 *Mathematics Subject Classification.* 65K05, 90C25, 90C30.
Key words and phrases. *self-adaptive technique, forward-backward algorithm, convex minimization problem, classification problem.*
Corresponding author: Prasit Cholamjiak; prasit.ch@up.ac.th

For any $\eta > 0$, it is well known that $\bar{e}$ is a solution of (1.1) if and only if it satisfies the following fixed point equation:

$$\bar{e} = \operatorname{prox}_{\eta G}(\bar{e} - \eta \nabla F(\bar{e})).$$

Here, the proximal operator $\operatorname{prox}_G : \mathbb{H} \to \operatorname{dom} G$ is defined by

$$\operatorname{prox}_G(e) = (Id + \partial G)^{-1}(e), \quad e \in \mathbb{H},$$

where Id denotes the identity operator on $\mathbb{H}$, ∇F is the gradient of F, and ∂G is the subdifferential of G, given by

$$\partial G(e) = \{b \in \mathbb{H} : G(d) - G(e) - \langle b, d - e\rangle \geq 0,\ \forall d \in \mathbb{H}\}.$$

Based on this fixed point formulation, the classical forward–backward splitting algorithm can be naturally derived as an iterative scheme that alternates between a forward (gradient) step and a backward (proximal) step. Specifically, the forward–backward algorithm is defined by

$$e_{n+1} = \underbrace{\operatorname{prox}_{\eta G}}_{\text{backward step}} \underbrace{\big(e_n - \eta \nabla F(e_n)\big)}_{\text{forward step}}, \quad n \geq 1.$$

The convergence analysis of the above iterative scheme relies on several fundamental properties of the proximal operator and auxiliary results from nonlinear analysis. In particular, the following characterization of the proximal operator will play a crucial role in establishing the convergence behavior of the proposed algorithms. For any $e \in \mathbb{H}$ and $\eta > 0$, it holds that

$$\frac{e - \operatorname{prox}_{\eta G}(e)}{\eta} \in \partial G(\operatorname{prox}_{\eta G}(e)).$$

This relation provides a direct link between the proximal mapping and the subdifferential of G, which is essential for deriving key inequalities in the subsequent analysis.

In 2000, Tseng [21] proposed the forward-backward-forward splitting algorithm, also known as Tseng's extragradient method. Starting from an initial point $e_0 \in \mathbb{H}$, the iterative scheme is given by

$$e_{n+1} = \operatorname{prox}_{\eta_n G}(e_n - \eta_n \nabla F(e_n)) - \eta_n\big(\nabla F(\operatorname{prox}_{\eta_n G}(e_n - \eta_n \nabla F(e_n))) - \nabla F(e_n)\big), \tag{1.2}$$

where (η_n) is a sequence of positive real numbers.

In 2005, Combettes and Wajs [5] introduced a relaxed variant of the forward–backward splitting method. Given an initial point $e_0 \in \mathbb{H}$ and $\varepsilon \in (0, \min\{1, 1/L\})$, the iteration is defined as

$$e_{n+1} = e_n + \lambda_n\big(\operatorname{prox}_{\eta_n G}(e_n - \eta_n \nabla F(e_n)) - e_n\big), \tag{1.3}$$

where $\lambda_n \in [\varepsilon, 1]$, $\eta_n \in [\varepsilon, (2/L) - \varepsilon]$, and L denotes the Lipschitz constant of ∇F.

Within the framework of convex optimization, Halpern-type iterative schemes [8] can be naturally incorporated into proximal gradient methods. Let F be a convex differentiable function with Lipschitz continuous gradient and let G be a proper lower semicontinuous convex function. We consider the convex minimization problem (1.1). A Halpern-type proximal gradient scheme [6, 16, 22] generates a sequence (e_n) according to the following iteration:

$$e_{n+1} = \kappa_n u + (1 - \kappa_n) \operatorname{prox}_{\eta_n G}\big(e_n - \eta_n \nabla F(e_n)\big),$$

where $u \in \mathbb{H}$ is a given reference point, $\eta_n > 0$ is a step size, and (κ_n) is a control sequence satisfying conditions analogous to those of the classical Halpern iteration. This scheme retains the structure of the Halpern iteration while incorporating the proximal operator to effectively handle nonsmooth components.

Furthermore, to facilitate the proof of weak and strong convergence results, we recall several technical lemmas that will be used throughout the paper. These results include boundedness criteria for iterative sequences, convergence principles for summable perturbations, and tools for weak convergence analysis such as Opial's condition, as well as demiclosedness properties of maximal monotone operators.

Next, we recall several technical lemmas that are essential for proving the main convergence results.

Lemma 1.1. [17] *Let $\Omega \subset \mathbb{H}$ and let $(e_n) \subset \mathbb{H}$ satisfy*

(1) for each $e \in \Omega$, the limit $\lim_{n\to\infty} \|e_n - e\|$ exists;

(2) $\omega_w(e_n) \subset \Omega$.

Then (e_n) converges weakly to a point in Ω.

Lemma 1.2. [10] *Let $\{a_n\}$ be a nonnegative sequence satisfying*

$$a_{n+1} \leq (1 - r_n)a_n + r_n q_n \quad \text{and} \quad a_{n+1} \leq a_n - b_n + p_n,$$

where $(r_n) \subset (0, 1)$, (b_n) is a nonnegative sequence, and $(p_n), (q_n)$ are real sequences such that

(1) $\sum_{n=1}^{\infty} r_n = \infty$;

(2) $\lim_{n\to\infty} p_n = 0$;

(3) $\lim_{j\to\infty} b_{n_j} = 0$ implies $\limsup_{j\to\infty} q_{n_j} \leq 0$ for any subsequence (n_j) of (n).

Then, $\lim_{n\to\infty} a_n = 0$.

Lemma 1.3. [1] *Let ∂G be a maximal monotone operator. Assume that*

$$\mathrm{Graph}(\partial G) = \{(e, a) \in \mathbb{H} \times \mathbb{H} : a \in \partial G(e)\}$$

is demiclosed, that is, if a sequence $(e_n, a_n) \subset \mathrm{Graph}(\partial G)$ satisfies that $e_n \rightharpoonup e$ and $a_n \to a$, then $(e, a) \in \mathrm{Graph}(\partial G)$.

These results form the analytical foundation for proving the main convergence theorems presented in the next section.

Highlights of this research:

- This paper proposes a novel self-adaptive forward–backward algorithm for solving convex minimization problems. The proposed scheme is developed within a splitting framework and is shown to achieve both weak and strong convergence under suitable assumptions.
- To validate the theoretical findings, numerical experiments are conducted in infinite-dimensional Hilbert spaces, demonstrating the stability and effectiveness of the proposed algorithm under various initial conditions.
- The applicability of the proposed method is further illustrated through data classification tasks. In particular, Section 3 presents its implementation for dementia prediction using a publicly available healthcare dataset from Kaggle (https://www.kaggle.com/datasets/timothyadeyemi/dementia-patient-health-dataset).
- Experimental results show that the proposed algorithm outperforms existing methods in terms of accuracy, precision, recall, and F1-score. Moreover, it achieves lower loss values and exhibits strong generalization capability across training and testing datasets.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed self-adaptive forward-backward algorithm and establish its weak and strong convergence under appropriate conditions. Section 3 is devoted to numerical experiments,

including examples in infinite-dimensional Hilbert spaces and real-world data classification, illustrating the efficiency and robustness of the proposed method in comparison with existing approaches. Finally, Section 4 summarizes the main results and discusses the effectiveness of the proposed algorithm.

2. Main Theorem

In this section, we propose a new forward-backward algorithm and study its convergence behavior. We begin by establishing a weak convergence theorem for the generated sequence and subsequently derive a strong convergence result by incorporating additional techniques.

First, we impose the following assumptions. Let $F, G : \mathbb{H} \to \mathbb{R} \cup (+\infty)$ be proper, lower semicontinuous, and convex. Assume that F is differentiable with Lipschitz continuous gradient ∇F on $\mathbb{H}$. Moreover, we assume that $\operatorname{argmin}(F+G) \neq \emptyset$.

Algorithm 2.1. (Adaptive Forward–Backward Splitting Algorithm (Weak Version))
Initialization: Given $\mu \in (0,1),\ \sigma \in (0,2)$ and $\eta_1 > 0$.
Iterative step: Let $e_1 \in \mathbb{H}$ and calculate e_{n+1} as follows:
Step1. Compute the forward-backward step:

$$d_n = \operatorname{prox}_{\eta_n G}(e_n - \eta_n \nabla F(e_n))$$

where

$$\eta_{n+1} = \begin{cases} \min\left\{ \dfrac{\mu(1+\ell_n)\|d_n - e_n\|}{\|\nabla F(d_n) - \nabla F(e_n)\|},\ (1+\ell_n)\eta_n + \delta_n \right\} & \text{if } \|\nabla F(d_n) - \nabla F(e_n)\| \neq 0 \\ (1+\ell_n)\eta_n + \delta_n & \text{otherwise.} \end{cases} \tag{2.1}$$

Step2. Set $\omega_n = \mu(1+\ell_n)$ and compute

$$e_{n+1} = e_n - \sigma\theta_n \mathrm{E}(e_n, d_n)$$

where $\mathrm{E}(e_n, d_n) = e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))$ and $\theta_n = \dfrac{(1-\omega_n)\|d_n - e_n\|^2}{\|\mathrm{E}(e_n, d_n)\|^2}$. Set $n = n+1$ and go to **Step1.**

Lemma 2.2. [14] *Let $\{\eta_n\}$ be defined by* (2.1)*. Suppose that (ℓ_n) and (δ_n) are nonnegative sequence such that $\sum_{n=1}^{\infty} \ell_n < \infty$ and $\sum_{n=1}^{\infty} \delta_n < \infty$. Then*

$$\lim_{n\to\infty} \eta_n = \eta \in \left[\min\left\{ \eta_1, \frac{\mu}{L} \right\}, e^a(\eta_1 + b) \right],$$

where $\sum_{n=1}^{\infty} \ell_n = a$ and $\sum_{n=1}^{\infty} \delta_n = b$. Moreover,

$$\|\nabla F(d_n) - \nabla F(e_n)\| \leq \frac{(1+\ell_n)\mu\|d_n - e_n\|}{\eta_{n+1}}, \quad n \geq 1.$$

Theorem 2.3. Let (e_n) be defined by Algorithm 2.1. Suppose that (ℓ_n) and (δ_n) are nonnegative sequences such that $\sum_{n=1}^{\infty} \ell_n < \infty$ and $\sum_{n=1}^{\infty} \delta_n < \infty$. Then, (e_n) weakly converges to an element of $\operatorname{argmin}(F+G)$.

Proof. Let e_* be a solution in $\operatorname{argmin}(F+G)$. Then, we obtain

$$\begin{aligned} \|e_{n+1} - e_*\|^2 &= \|e_n - \sigma\theta_n \mathrm{E}(e_n, d_n) - e_*\|^2 \\ &= \|e_n - e_*\|^2 - 2\sigma\theta_n \langle e_n - e_*, \mathrm{E}(e_n, d_n)\rangle + \sigma^2\theta_n^2 \|\mathrm{E}(e_n, d_n)\|^2. \end{aligned} \tag{2.2}$$

Using the definition of $\mathrm{E}(e_n, d_n)$ and (2.1), we have

$$\begin{aligned}
\langle e_n - e_*, \mathrm{E}(e_n, d_n)\rangle &= \langle e_n - d_n, \mathrm{E}(e_n, d_n)\rangle + \langle d_n - e_*, \mathrm{E}(e_n, d_n)\rangle \\
&= \langle e_n - d_n, e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle \\
&\quad + \langle d_n - e_*, e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle \\
&= \|e_n - d_n\|^2 - \langle e_n - d_n, \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle \\
&\quad + \langle d_n - e_*, e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle \\
&\geq \|e_n - d_n\|^2 - \eta_n\|e_n - d_n\|\|\nabla F(e_n) - \nabla F(d_n)\| \\
&\quad + \langle d_n - e_*, e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle \\
&\geq \|e_n - d_n\|^2 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\|e_n - d_n\|^2 \\
&\quad + \langle d_n - e_*, e_n - d_n - \eta_n(\nabla F(e_n) - \nabla F(d_n))\rangle. \qquad (2.3)
\end{aligned}$$

Since $d_n = \mathrm{prox}_{\eta_n G}(e_n - \eta_n \nabla F(e_n))$, it follows that $(\mathrm{I} - \eta_n\nabla F)e_n \in (\mathrm{I} - \eta_n \partial G)d_n$. Moreover, we have $(\mathrm{I} - \eta_n\nabla F)e_n = d_n - \eta_n c_n$ for some $c_n \in \partial G(d_n)$ This shows that

$$c_n = \frac{1}{\eta_n}\big(e_n - d_n - \eta_n \nabla F(e_n)\big). \qquad (2.4)$$

On the other hand, we get $0 \in \nabla F(e_*) + \partial G(e_*) \subseteq \partial(F+G)(e_*)$ and $\nabla F(d_n) + c_n \in \partial(F+G)(d_n)$. Therefore, by the maximality of $\partial(F+G)$, we obtain

$$\langle \nabla F(d_n) + c_n, d_n - e_*\rangle \geq 0. \qquad (2.5)$$

Substitution of (2.4) into (2.5) yields

$$\big\langle \nabla F(d_n) + \frac{1}{\eta_n}\big(e_n - d_n - \eta_n\nabla F(e_n)\big), d_n - e_*\big\rangle \geq 0.$$

This gives

$$\langle e_n - d_n - \eta_n\nabla F(e_n) + \eta_n \nabla F(d_n), d_n - e_*\rangle \geq 0. \qquad (2.6)$$

By (2.3) and (2.6), we see that

$$\begin{aligned}
\langle e_n - e_*, \mathrm{E}(e_n, d_n)\rangle &\geq \|e_n - d_n\|^2 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\|e_n - d_n\|^2 \\
&= \left(1 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right)\|e_n - d_n\|^2. \qquad (2.7)
\end{aligned}$$

From (2.2) and (2.7), we obtain

$$\begin{aligned}
\|e_{n+1} - e_*\|^2 &\leq \|e_n - e_*\|^2 - 2\sigma\theta_n\left(1 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right)\|e_n - d_n\|^2 + \sigma^2\theta_n^2\|E(e_n, d_n)\|^2 \\
&= \|e_n - e_*\|^2 - 2\sigma\theta_n\left(1 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right)\|e_n - d_n\|^2 \\
&\quad + \sigma^2\theta_n(1 - \mu(1+\ell_n))\|e_n - d_n\|^2 \\
&= \|e_n - e_*\|^2 - \left[2\sigma\theta_n\left(1 - \frac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right) - \sigma^2\theta_n(1 - \mu(1+\ell_n))\right]\|e_n - d_n\|^2. \qquad (2.8)
\end{aligned}$$

Set $t_n = 2\sigma\theta_n\left(1 - \dfrac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right) - \sigma^2\theta_n(1-\mu(1+\ell_n))$. Moreover, we see that

$$\begin{aligned}\|\mathrm{E}(e_n,d_n)\|^2 &= \|e_n-d_n\|^2 + \eta_n^2\|\nabla F(e_n)-\nabla F(d_n)\|^2 - 2\eta_n\langle e_n-d_n, \nabla F(e_n)-\nabla F(d_n)\rangle\\ &\leq \|e_n-d_n\|^2 + \frac{\eta_n^2\mu^2(1+\ell_n)}{\eta_{n+1}^2}\|e_n-d_n\|^2\\ &= \left(1+\frac{\eta_n^2\mu^2(1+\ell_n)^2}{\eta_{n+1}^2}\right)\|e_n-d_n\|^2.\end{aligned}$$

It follows that

$$\begin{aligned}\theta_n &= \frac{(1-\mu(1+\ell_n))\|e_n-d_n\|^2}{\|E_n(e_n,d_n)\|^2}\\ &\geq \frac{(1-\mu(1+\ell_n))}{s_n}, \text{ where } s_n = 1+\frac{\eta_n^2\mu^2(1+\ell_n)^2}{\eta_{n+1}^2}, n\geq 1.\end{aligned}$$

From Lemma 2.2 and $\displaystyle\sum_{n=1}^{\infty}\ell_n < \infty$, we see that there is $N\in\mathbb{N}$ such that

$$\theta_n > \frac{1-\mu}{1+\mu^2} > 0 \text{ for all } n\geq N.$$

Moreover, we have, for each $n\geq N$,

$$\begin{aligned}t_n &> \frac{2\sigma(1-\mu)^2}{1+\mu^2} - \frac{\sigma^2(1-\mu)^2}{1+\mu^2}\\ &= \frac{(2-\sigma)\sigma(1-\mu)^2}{1+\mu^2}\\ &> 0.\end{aligned}$$

So, for each $n\geq N$, we obtain

$$\|e_{n+1}-e_*\|^2 \leq \|e_n-e_*\|^2 - \frac{(2-\sigma)\sigma(1-\mu)^2}{1+\mu^2}\|e_n-d_n\|^2. \tag{2.9}$$

So $(\|e_n-e_*\|)$ is a convergent sequence and it is also a bounded sequence. Hence, (e_n) is bounded. From (2.9), we get $\|e_n-d_n\|\to 0$ as $n\to\infty$. By the boundedness of (e_n), we know that the set of its weak accumulation points is nonempty. Let e_∞ be a weak accumulation point of (e_n). Therefore, there is subsequence (e_{n_i}) of (e_n). Next, we show that $e_\infty\in\operatorname{argmin}(F+G)$. Let $(a,b)\in\operatorname{Graph}(\nabla F+\partial G)$, that is $b-\nabla(a)\in\partial G(a)$. Since $d_{n_i} = \operatorname{prox}_{\eta_{n_i}G}(e_{n_i}-\eta_{n_i}\nabla F(e_{n_i}))$, we obtain

$$(\mathrm{I}-\eta_{n_i}\nabla F)e_{n_i}\in(\mathrm{I}+\eta_{n_i}\partial G)d_{n_i}.$$

This yields $\dfrac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}-\eta_{n_i}\nabla F(e_{n_i})\big)\in\partial G(d_{n_i})$. Since ∂G is maximal monotone, we have

$$\left\langle a-d_{n_i}, b-\nabla F(a)-\frac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}-\eta_{n_i}\nabla F(e_{n_i})\big)\right\rangle \geq 0.$$

This shows that

$$\begin{aligned}\langle a-d_{n_i},b\rangle &\geq \big\langle a-d_{n_i},\nabla F(a)+\frac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}-\eta_{n_i}\nabla F(e_{n_i})\big)\big\rangle\\ &= \langle a-d_{n_i},\nabla F(a)-\nabla F(e_{n_i})\rangle+\big\langle a-d_{n_i},\frac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}\big)\big\rangle\\ &= \langle a-d_{n_i},\nabla F(a)-\nabla F(d_{n_i})\rangle+\langle a-d_{n_i},\nabla F(d_{n_i})-\nabla F(e_{n_i})\rangle\\ &\quad+\big\langle a-d_{n_i},\frac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}\big)\big\rangle\\ &\geq \langle a-d_{n_i},\nabla F(d_{n_i})-\nabla F(e_{n_i})\rangle+\big\langle a-d_{n_i},\frac{1}{\eta_{n_i}}\big(e_{n_i}-d_{n_i}\big)\big\rangle.\end{aligned}$$

Since $\lim_{i\to\infty}\|e_{n_i}-d_{n_i}\|=0$ and F is differentiable with Lipschitz continuous gradient ∇F on $\mathbb{H}$, we have $\lim_{i\to\infty}\|\nabla F(d_{n_i})-\nabla F(e_{n_i})\|=0$. Hence, we obtain

$$\langle a-e_\infty,b\rangle=\lim_{i\to\infty}\langle a-d_{n_i},b\rangle\geq 0.$$

This shows that $0\in \operatorname{argmin}(F+G)e_\infty$, and consequently $e_\infty\in\operatorname{argmin}(F+G)$ by Lemma 1.3. This confirms that (e_n) converges weakly to an element of $\operatorname{argmin}(F+G)$ by Lemma 1.1. We thus complete the proof. □

Next, we establish a strong convergence theorem for the sequence generated by the proposed algorithm. To this end, we incorporate the Halpern-type iterative technique into our scheme, which enables us to strengthen the weak convergence result obtained earlier.

Algorithm 2.4. (Adaptive Forward–Backward Splitting Algorithm (Strong Version))
Initialization: Given $\mu\in(0,1),\ \sigma\in(0,2)$ and $\eta_1>0$.
Iterative step: Let $e_1,u\in\mathbb{H}$ and calculate e_{n+1} as follows:
Step1. Compute the forward-backward step:

$$d_n=\operatorname{prox}_{\eta_n G}(e_n-\eta_n\nabla F(e_n))$$

where

$$\eta_n(2.10)\begin{cases}\min\big\{\dfrac{\mu(1+\ell_n)\|d_n-e_n\|}{\|\nabla F(d_n)-\nabla F(e_n)\|},\ (1+\ell_n)\eta_n+\delta_n\big\} & \text{if } \|\nabla F(d_n)-\nabla F(e_n)\|\neq 0\\ (1+\ell_n)\eta_n+\delta_n & \text{otherwise.}\end{cases}$$

Step2. Set $\omega_n=\mu(1+\ell_n)$ and compute

$$e_{n+1}=\kappa_n u+(1-\kappa_n)[e_n-\sigma\theta_n\mathrm{E}(e_n,d_n)]$$

where $\mathrm{E}(e_n,d_n)=e_n-d_n-\eta_n(\nabla F(e_n)-\nabla F(d_n))$ and $\theta_n=\dfrac{(1-\omega_n)\|d_n-e_n\|^2}{\|\mathrm{E}(e_n,d_n)\|^2}$. Set $n=n+1$ and go to **Step1.**

Theorem 2.5. Let (e_n) be defined by Algorithm 2.4. Suppose that (κ_n), (ℓ_n) and (δ_n) are nonnegative sequences such that $\lim_{n\to\infty}\kappa_n=0$, $\sum_{n=1}^{\infty}\kappa_n=\infty$, $\sum_{n=1}^{\infty}\ell_n<\infty$ and $\sum_{n=1}^{\infty}\delta_n<\infty$. Then, (e_n) strongly converges to an element of $\operatorname{argmin}(F+G)$.

Proof. Let $\bar{e}=\mathrm{P}_{\operatorname{argmin}(F+G)}(u)\in\operatorname{argmin}(F+G)$. Then, we obtain

$$\begin{aligned}\|e_{n+1}-\bar{e}\|^2 &= \|\kappa_n u+(1-\kappa_n)[e_n-\sigma\theta_n\mathrm{E}(e_n,d_n)]-\bar{e}\|^2\\ &\leq (1-\kappa_n)\|e_n-\sigma\theta_n\mathrm{E}(e_n,d_n)-\bar{e}\|^2+2\kappa_n\langle u-\bar{e},e_{n+1}-\bar{e}\rangle.\end{aligned}\tag{2.11}$$

Using the same proof as in Theorem 2.3, we obtain

$$\|e_n - \sigma\theta_n \mathrm{E}(e_n, d_n) - \bar{e}\|^2 \leq \|e_n - \bar{e}\|^2 - t_n\|e_n - d_n\|^2$$

where $t_n = 2\sigma\theta_n\left(1 - \dfrac{\eta_n\mu(1+\ell_n)}{\eta_{n+1}}\right) - \sigma^2\theta_n(1-\mu(1+\ell_n))$. Nothing that there exists $N \in \mathbb{N}$ such that

$$t_n > \frac{(2-\sigma)\sigma(1-\mu)^2}{1+\mu^2} > 0, \quad \text{for all } n \geq N.$$

Using (2.11), we see that

$$\begin{aligned}\|e_{n+1} - \bar{e}\|^2 &\leq (1-\kappa_n)[\|e_n - \bar{e}\|^2 - t_n\|e_n - d_n\|^2] + 2\kappa_n\langle u - \bar{e}, e_{n+1} - \bar{e}\rangle \\ &= (1-\kappa_n)\|e_n - \bar{e}\|^2 - t_n(1-\kappa_n)\|e_n - d_n\|^2 + 2\kappa_n\langle u - \bar{e}, e_{n+1} - \bar{e}\rangle.\end{aligned}$$

Moreover, for each $n \geq N$, we have

$$\begin{aligned}\|e_{n+1} - \bar{e}\| &\leq \kappa_n\|u - \bar{e}\| + (1-\kappa_n)\|(e_n - \bar{e}) - \sigma\theta_n \mathrm{E}(e_n, d_n)\| \\ &\leq \kappa_n\|u - \bar{e}\| + (1-\kappa_n)\|e_n - \bar{e}\| \\ &\leq \max\{\|u - \bar{e}\|, \|e_n - \bar{e}\|\} \\ &\vdots \\ &\leq \max\{\|u - \bar{e}\|, \|e_N - \bar{e}\|\}.\end{aligned}$$

This shows that (e_n) is boundad. To invoke Lemma 1.2, we define

$$\begin{aligned}p_n &= 2\kappa_n\langle u - \bar{e}, e_{n+1} - \bar{e}\rangle, \\ b_n &= t_n(1-\kappa_n)\|e_n - d_n\|, \\ r_n &= \kappa_n, \\ q_n &= 2\langle u - \bar{e}, e_{n+1} - \bar{e}\rangle, \\ a_n &= \|e_n - \bar{e}\|^2.\end{aligned}$$

We can check that conditions *(1)* and *(2)* in Lemma 1.2 are satisfied. Next, we will show that condition *(3)* holds. Let a subsequence of (n) be (n_i) and suppose that $\lim_{i\to\infty} \mathrm{B}_{n_i} = 0$. Then, we get $\lim_{i\to\infty} \|e_{n_i} - d_{n_i}\| = 0$. Since (e_{n_i}) is bounded, it admits at least one weak accumulation point. Let $\bar{\bar{e}}$ be such a point. Then, there exists a subsequence $(e_{n_{i_k}})$ that weakly converges to $\bar{\bar{e}}$. Moreover, $(e_{n_{i_k}})$ also weakly converges to the same point. Since $\lim_{k\to\infty} \|d_{n_{i_k}} - e_{n_{i_k}}\| = 0$ and ∇F is Lipschitz continuous, we obtain

$$\lim_{k\to\infty} \|\nabla F(d_{n_{i_k}}) - \nabla F(e_{n_{i_k}})\| = 0.$$

By definition of $(d_{n_{i_k}})$, we see that

$$\frac{1}{\eta_{n_{i_k}}}\left(e_{n_{i_k}} - d_{n_{i_k}} - \eta_{n_{i_k}}\nabla F(e_{n_{i_k}})\right) \in \partial G(d_{n_{i_k}}),$$

which implies that

$$\begin{aligned}\frac{e_{n_{i_k}} - d_{n_{i_k}}}{\eta_{n_{i_k}}} + \nabla F(d_{n_{i_k}}) - \nabla F(e_{n_{i_k}}) &\in \nabla F(d_{n_{i_k}}) + \partial G(d_{n_{i_k}}) \\ &\subseteq \partial(F+G)(d_{n_{i_k}}).\end{aligned}$$

Passing $k \to \infty$, by Lemma 1.3, we conclude that $\bar{\bar{e}} \in \operatorname{argmain}(F+G)$. It follows that

$$\begin{aligned}\limsup_{i\to\infty}\langle u-\bar{e}, e_{n_i}-\bar{e}\rangle &= \lim_{k\to\infty}\langle u-\bar{e}, e_{n_{i_k}}-\bar{e}\rangle \\ &= \langle u-\bar{e}, \bar{\bar{e}}-\bar{e}\rangle \\ &\leq 0.\end{aligned}$$

Moreover, we see that

$$\begin{aligned}\|e_{n+1}-e_n\| &\leq \kappa_n\|u-e_n\| + (1-\kappa_n)\sigma\theta_n\|\mathrm{E}(e_n,d_n)\| \\ &\leq \kappa_n\|u-e_n\| + (1-\kappa_n)\sigma\theta_n(\|e_n-d_n\| + \eta_n\|\nabla F(e_n)-\nabla F(d_n)\|).\end{aligned}$$

So, we have $\|e_{n_i}-e_{n_i+1}\| \to 0$ as $i\to\infty$. Thus, we obtain $\limsup_{i\to\infty}\langle u-\bar{e}, e_{n_i+1}-\bar{e}\rangle \leq 0$. Hence, we get $\limsup_{i\to\infty} \mathrm{Q}_{n_i} \leq 0$. We thus complete the proof. □

3. Numerical Experiments

In this section, we conduct numerical experiments to illustrate the performance of the proposed algorithms. The experiments are divided into two parts. In the first part, we consider an example in an infinite-dimensional Hilbert space to demonstrate the strong convergence behavior of the iterative scheme corresponding to Algorithm 2.4. This example is designed to support the theoretical results and to examine the effect of the inertial term on both convergence speed and stability. In the second part, we apply the proposed method to a data classification problem using the Dementia Patient Health Dataset in order to illustrate the weak convergence behavior associated with Algorithm 2.1. This experiment further demonstrates the applicability of the algorithm in practical situations, particularly in medical data analysis. All numerical computations are carried out using MATLAB 2022b on a 64-bit MacBook Pro with an Apple M1 chip and 8 GB of RAM.

3.1. **Example in infinite dimensional space.**

In this subsection, we implement one of the proposed iterative schemes (Algorithm 2.4) to solve an optimization problem in an infinite-dimensional Hilbert space. The purpose of this experiment is to examine the convergence behavior and numerical performance of the algorithm under a setting where strong convergence can be theoretically ensured. In particular, we investigate the role of the inertial parameter incorporated into the iteration process, with emphasis on its influence on convergence acceleration and stability.

Let $L_2[0,1] = \{a(t) : \int_0^1 |a(t)|^2 \mathrm{d}t < \infty\}$ be a Hilbert space equipped with the norm

$$\|a\| = \sqrt{\int_0^1 |a(t)|^2 \mathrm{d}t}$$

and the inner product defined by

$$\langle a, b\rangle = \int_0^1 a(t)b(t)\mathrm{d}t, \text{ for all } b(t) \in L_2[0,1].$$

We consider the following minimization problem:

$$\min_{a(t)\in L_2[0,1]} (2\langle t, a(t)\rangle + 2\|a\|).$$

Define the functions F and G by

$$F(a(t)) = 2\langle t, a(t)\rangle, \quad G(a(t)) = 2\|a\|.$$

It is clear that f is convex, proper, and lower semicontinuous. Moreover, its gradient $\nabla F(a(t)) = 2a(t)$ is Lipschitz continuous with 2-Lipschitz constant.

The proximal operator of G is given by

$$\text{prox}_{\eta_n G}(a(t)) = \begin{cases} \left(1 - \dfrac{2\eta_n}{\|a\|}\right) a(t) & \text{if } \|a\| > 2\eta_n, \\ 0 & \text{otherwise.} \end{cases}$$

By using Algorithm 2.4, we approximate the solution to the minimization problem. The approximation takes into account the initialization settings where a_1 belongs to $L^2[0,1]$. The stopping criterion is defined as:

$$\mathrm{E}_n = \|a_n(s) - a_{n+1}(s)\|_{L^2}^2 \leq 10^{-K}.$$

We consider five different initial points e_1 as shown in the upper-right panel of Figure 1 and Figure 2. The parameters are chosen as

$$\mu = 0.01, \quad \sigma = 0.5, \quad \eta_0 = 10^{-5}, \quad \ell_n = \frac{1}{(1+n)^2}, \quad \delta_n = \frac{1}{(2n+2)^4}, \quad \kappa_n = \frac{1}{(8n+17)^{0.91}}.$$

The results are reported in Tables 1-2 and Figures 1-2. We denote by CPU the computational time, and by Iter the number of iterations. The results show that the proposed algorithm converges for all initial points e_0.

TABLE 1. Numerical results for different initial points e_0 with $K = 3$.

Initial point e_0	Iter	Time	$\text{Error}_{1:5}$	$\text{Error}_{\text{final}}$
$\cos(t) - 7$	43	7.2307	[2.5058, 0.1239, 0.0762, 0.0519, 0.0378]	9.7513×10^{-4}
$16t^3 - 1$	30	4.7111	[0.2752, 0.0098, 0.0203, 0.0231, 0.0205]	9.5493×10^{-4}
$6t^3$	18	2.7569	[0.1044, 0.0087, 0.0082, 0.0088, 0.0078]	9.0628×10^{-4}
$12\sin(t) - 1$	29	4.5069	[0.2572, 0.0228, 0.0247, 0.0226, 0.0188]	9.5223×10^{-4}
$11 - \cos(t)$	41	6.4154	[0.4983, 0.0618, 0.0536, 0.0448, 0.0359]	9.8147×10^{-4}

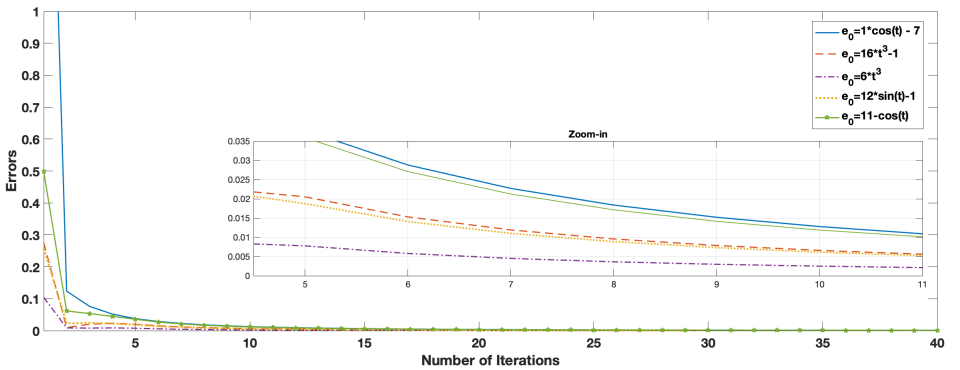


FIGURE 1. The graph of error and number of iterations

From Tables 1-2 and Figures 1-2, it shows that the proposed algorithm converges for all initial points e_0 under both choices of K. It can be observed that, for $K = 3$, the algorithm requires fewer iterations and less computational time, while the obtained errors are of order 10^{-4}. In contrast, for $K = 4$, the algorithm achieves higher accuracy with errors of order 10^{-5}, at the expense of increased iterations and computational time. Moreover, the convergence behavior remains consistent across all initial points, indicating that the method is stable and robust with respect to different initializations.

TABLE 2. Numerical results for different initial points e_0 with $K = 4$.

Initial point e_0	Iter	Time	$\text{Error}_{1:5}$	$\text{Error}_{\text{final}}$
$\cos(t) - 7$	146	23.4721	$[2.5058,\ 0.1239,\ 0.0762,\ 0.0519,\ 0.0378]$	9.9167×10^{-5}
$16t^3 - 1$	102	15.5780	$[0.2752,\ 0.0098,\ 0.0203,\ 0.0231,\ 0.0205]$	9.8489×10^{-5}
$6t^3$	61	9.1522	$[0.1044,\ 0.0087,\ 0.0082,\ 0.0088,\ 0.0078]$	9.7863×10^{-5}
$12\sin(t) - 1$	98	14.9431	$[0.2572,\ 0.0228,\ 0.0247,\ 0.0226,\ 0.0188]$	9.9722×10^{-5}
$11 - \cos(t)$	140	21.6020	$[0.4983,\ 0.0618,\ 0.0536,\ 0.0448,\ 0.0359]$	9.8890×10^{-5}

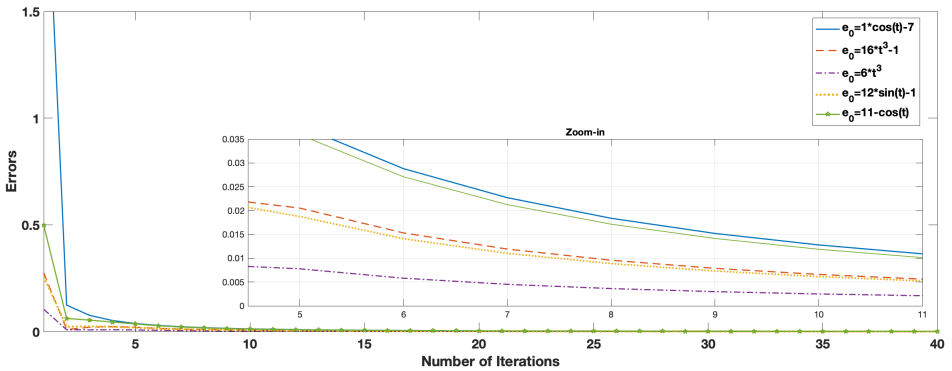


FIGURE 2. The graph of error and number of iterations

3.2. **Data Classification.** In recent years, the number of elderly individuals has been steadily increasing, leading to a higher risk of cognitive decline and dementia-related conditions. This growing concern highlights the importance of early detection and effective screening methods. For this reason, we employ the Dementia Patient Health Dataset as a benchmark to evaluate the performance of the proposed algorithm. This dataset provides relevant health and behavioral features associated with cognitive conditions, making it suitable for classification tasks. Furthermore, the insights gained from this study are expected to support future applications of the proposed method to real-world data collected in practical settings.

The dataset used in this study is obtained from the publicly available Dementia Patient Health Dataset provided on Kaggle (https://www.kaggle.com/datasets/timothyadeyemi/dementia-patient-health-dataset). This dataset consists of individual-level health information, including demographic data, physiological measurements, lifestyle factors, and clinical-related variables associated with cognitive conditions. In total, the dataset contains 24 features, which can be categorized into several groups according to their characteristics. These features include numerical variables (e.g., age, heart rate, and body temperature), categorical variables (e.g., gender, smoking status, and education level), and clinical indicators (e.g., cognitive test scores and medical history). For this study, the data are analyzed through a classification framework. Each feature is systematically organized as shown in Table 3.

Tables 4 and 5 present the descriptive analysis of the dataset, including both numerical and categorical features. Table 4 summarizes the statistical characteristics of numerical variables in terms of mean, standard deviation, and range (minimum and maximum values). It can be observed that the age of participants ranges from 60 to 90 years, with an

TABLE 3. Description of dataset features

Attribute	Description	Value type	Variable type
Age	Age of patient	int (years)	Demographic
Gender	Patient's gender	binary (0,1)	Demographic
Education_Level	Education level	categorical	Demographic
Weight	Body weight	float (kg)	Clinical
HeartRate	Heart rate	int (bpm)	Clinical
BloodOxygenLevel	Oxygen saturation	float (%)	Clinical
BodyTemperature	Body temperature	float (°C)	Clinical
Diabetic	Diabetes status	binary	Medical history
Smoking_Status	Smoking behavior	categorical	Lifestyle
AlcoholLevel	Alcohol consumption	categorical	Lifestyle
Physical_Activity	Activity level	categorical	Lifestyle
Sleep_Quality	Sleep quality	categorical	Lifestyle
Nutrition_Diet	Dietary pattern	categorical	Lifestyle
Depression_Status	Depression condition	binary	Psychological
Cognitive_Test_Scores	Cognitive score	int (1–10)	Clinical
Family_History	Family history of dementia	binary	Genetic
APOE_e4	Genetic risk factor	binary	Genetic
Medication_History	Medication usage	categorical	Medical history
Prescription	Type of prescription	categorical	Treatment
Dosage	Drug dosage	float (mg)	Treatment
MRI_Delay	MRI delay	int (days)	Clinical
Chronic_Health_Conditions	Chronic diseases	categorical	Medical history
Dominant_Hand	Dominant hand	categorical	Demographic
Dementia	Dementia status	binary (0,1)	Target

average of 74.91 years, indicating that the dataset primarily represents an elderly population. Similarly, clinical indicators such as heart rate, blood oxygen level, and body temperature fall within normal physiological ranges, suggesting that the dataset is well-structured and consistent with real-world medical data. Notably, the cognitive test scores exhibit a relatively wide variation (0 to 10), which may play a significant role in distinguishing between dementia and non-dementia cases.

TABLE 4. Descriptive statistics of numerical features

Feature	Mean $\pm$ Std	(Min, Max)
Age	74.91 ± 9.10	(60, 90)
Weight	74.32 ± 14.46	(50.07, 99.98)
HeartRate	79.38 ± 12.11	(60, 100)
BloodOxygenLevel	95.23 ± 2.93	(90.01, 100.00)
BodyTemperature	36.76 ± 0.43	(36.00, 37.50)
AlcoholLevel	0.098 ± 0.059	(0.0004, 0.1999)
MRI_Delay	30.10 ± 16.86	(0.09, 59.96)
Dosage	9.21 ± 6.49	(1.50, 23.00)
Cognitive_Test_Scores	6.38 ± 3.18	(0, 10)

*Mean $\pm$ Std denotes the average value and its standard deviation, indicating the central tendency and variability of the data.

Table 5 presents the distribution of categorical variables. The gender distribution is nearly balanced between female (50.4%) and male (49.6%), reducing potential bias related to gender. For lifestyle-related variables, smoking status shows that the majority of participants are either never-smokers or former smokers, while physical activity levels are relatively evenly distributed across mild, moderate, and sedentary groups. Importantly, the target variable (Dementia) is also balanced (51.5% vs. 48.5%), which is advantageous for classification tasks as it prevents model bias toward a dominant class. Overall, these descriptive statistics indicate that the dataset is well-balanced and suitable for developing and evaluating machine learning models for dementia classification.

TABLE 5. Frequency distribution of categorical variables

Feature	Category	Frequency (%)
Gender	Female / Male	50.4 / 49.6
Smoking_Status	Never / Former / Current	45.2 / 45.8 / 9.0
Physical_Activity	Mild / Moderate / Sedentary	35.1 / 31.8 / 33.1
Dementia	0 / 1	51.5 / 48.5

Data classification is a fundamental process for extracting meaningful patterns from datasets and has been widely applied across both scientific research and practical applications. In recent years, mathematically grounded algorithms, particularly those based on optimization theory and iterative methods, have been increasingly utilized to enhance classification performance. These approaches typically involve minimizing well-defined objective functions or solving variational problems, which enables accurate and stable separation of data into predefined categories, even in the presence of complex and high-dimensional structures.

In the medical domain, the application of mathematical classification techniques plays a crucial role in disease prediction, clinical decision support, and risk assessment. In particular, dementia is a complex public health problem associated with multiple factors, including demographic characteristics, health behaviors, and cognitive test scores. The use of mathematical approaches allows for systematic analysis of these interrelated variables and facilitates effective classification of individuals according to their risk levels, thereby supporting early screening and appropriate clinical management. In this study, we propose the application of an optimization-based algorithm originally developed for convex

minimization problems to the classification of dementia data. The dataset used in this study is the Dementia Patient Health Dataset, which consists of 1,000 individual health records, including features related to clinical conditions, behavioral patterns, and relevant risk factors. The dataset is partitioned into a training set comprising 80% (800 samples) and a testing set comprising 20% (200 samples), in order to systematically develop and evaluate the performance of the proposed model.

Next, we employ the Extreme Learning Machine (ELM) proposed by Huang et al. [11] as the core classification model due to its fast learning capability and strong generalization performance. ELM determines output weights analytically while randomly assigning input weights and hidden biases, which significantly reduces computational complexity. The ELM is integrated with the proposed optimization framework to improve classification accuracy for dementia prediction. The concept is introduced by defining the training set as follows:

Let the dataset be defined by

$$e := \{(e_i, t_i) : e_i \in \mathbb{R}^n,\ t_i \in \mathbb{R}^m,\ i = 1, 2, \ldots, N\},$$

where e_i denotes the input vector and t_i represents the corresponding target output.

Consider a single-hidden layer feedforward neural network (SLFN) with M hidden nodes. The network output associated with the i-th sample is expressed as

$$\mathbb{O}_i = \sum_{n=1}^{M} F_n\, U(w_n e_i + b_n),$$

where U is the activation function, w_n and b_n are the input weight matrix and bias vector of the n-th hidden node, respectively, and F_n denotes the output weight.

To characterize the hidden layer responses, define the matrix $\mathbb{L} \in \mathbb{R}^{P \times M}$ as

$$\mathbb{L} = \begin{bmatrix} U(w_1 e_1 + b_1) & \cdots & U(w_M e_1 + b_M) \\ \vdots & \ddots & \vdots \\ U(w_1 e_P + b_1) & \cdots & U(w_M e_P + b_M) \end{bmatrix}.$$

The ELM training procedure seeks a vector

$$F = [F_1^T, \ldots, F_M^T]^T$$

such that

$$\mathbb{L}F = \mathbb{S},$$

where

$$\mathbb{S} = [t_1^T, \ldots, t_P^T]^T$$

collects the desired outputs.

A common solution is obtained via the Moore–Penrose generalized inverse, given by

$$F = \mathbb{L}^{\dagger}\mathbb{S}.$$

Alternatively, the problem can be formulated as a regularized least-squares model

$$\min_{F \in \mathbb{R}^M} \left\{ \frac{1}{2} \|\mathbb{L}F - \mathbb{S}\|_2^2 + \lambda \|F\|_1 \right\},$$

where $\lambda > 0$ is a regularization parameter associated with the LASSO framework [20].

To evaluate the predictive capability of the proposed framework, four commonly adopted performance indices are employed, namely precision, recall, F1-measure, and accuracy. These quantities are expressed as

$$\textbf{Precision (PRE)} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\%,$$

$$\textbf{Recall (REC)} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\%,$$

$$\textbf{F1-score (F1)} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \times 100\%,$$

$$\textbf{Accuracy (ACC\%)} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \times 100\%.$$

Here, TP and TN correspond to correctly identified positive and negative samples, whereas FP and FN indicate incorrectly assigned positive and negative outcomes, respectively.

To further quantify the model performance, the binary cross-entropy function is adopted as a loss criterion. This function evaluates the deviation between predicted values and true labels, and is formulated as

$$\textbf{Loss} = -\frac{1}{\text{output size}} \sum_{i=1}^{\text{output size}} \left(e_i \log \bar{e}_i + (1-e_i)\log(1-\bar{e}_i)\right),$$

where the output size denotes the total number of scalar outputs, e_i represents the true label, and $\bar{e}_i$ is the predicted value corresponding to the i-th component.

For the comparison of our Algorithms 2.1 with Algorithm 1.3 [5] and Algorithm 1.2 [21]. The parameters are chosen as

$$\mu = 0.01, \quad \sigma = 0.5, \quad \eta_0 = 10^{-5}, \quad \ell_n = \frac{1}{(1+n)^2}, \quad \delta_n = \frac{1}{(2n+2)^4}, \quad \kappa_n = \frac{1}{(8n+17)^{0.91}}.$$

We choose $\beta_n = \dfrac{0.1}{\max(\mathrm{eig}A^TA)}$ in Algorithm 1.3 [5] and $\eta_n = \dfrac{0.1}{\max(\mathrm{eig}A^TA)}$ in Algorithm 1.3 [5] and Algorithm 1.2 [21].

The computational process is initialized by employing the sigmoid function as the activation function, while fixing the number of hidden nodes at $\mathbb{L} = 340$. The algorithm is executed under a predefined stopping criterion, where the iterative process is terminated once the maximum number of iterations (2500 iterations) is reached. The corresponding results are summarized in Tables 6 and 7.

TABLE 6. Performance comparison of different methods

Method	$\mathrm{Acc}_{\mathrm{train}}$	$\mathrm{Acc}_{\mathrm{test}}$	Precision	Recall	F1
OUR	93.2836	95.5000	92.7273	99.0291	95.7746
Combettes	92.1642	95.0000	91.8919	99.0291	95.3271
Tseng	90.1741	91.5000	87.0690	98.0583	92.2374

From Table 6, it can be observed that the proposed method consistently outperforms the Combettes and Tseng methods across all evaluation metrics. Specifically, the proposed method achieves the highest training accuracy (93.2836) and testing accuracy (95.5000), indicating superior generalization performance. In terms of precision, the proposed method also yields the best result (92.7273), reflecting its effectiveness in correctly identifying positive cases. Moreover, it maintains an exceptionally high recall (99.0291), which is crucial in medical screening applications to minimize missed diagnoses. Consequently, the proposed method attains the highest F1-score (95.7746), demonstrating a well-balanced trade-off between precision and recall. In comparison, although the Combettes method provides competitive results, it is slightly inferior in all metrics, while the Tseng method

shows noticeably lower performance. Overall, these findings confirm that the proposed method offers improved classification accuracy and reliability.

To further evaluate the effectiveness of the proposed approach, a comparative study with existing methods is conducted based on the binary cross-entropy loss. The results are presented in Table 7.

TABLE 7. Comparison of algorithm performance

Algorithms	Iterations	Time	Phase	Loss
Our method	2500	0.4417	Training	0.3555
			Testing	0.0875
Combettes	2500	0.8683	Training	0.4034
			Testing	0.0987
Tseng	2500	1.2962	Training	0.4375
			Testing	0.1075

From Table 7, It can be observed that the proposed method consistently achieves lower loss values compared to the Combettes and Tseng methods in both training and testing phases. In particular, the proposed method attains a training loss of 0.3555 and a testing loss of 0.0875, which are significantly lower than those of the Combettes method (0.4034 and 0.0987) and the Tseng method (0.4375 and 0.1075), respectively. This indicates that the proposed algorithm is more effective in minimizing the discrepancy between predicted outputs and true labels. Moreover, the proposed method requires less computational time, suggesting improved efficiency alongside enhanced predictive performance.

The results presented in Tables 6 and 7 are further illustrated in Figure 3, where the evolution of accuracy and loss over iterations can be observed. The graphical representation provides a clearer view of the convergence behavior of the model.

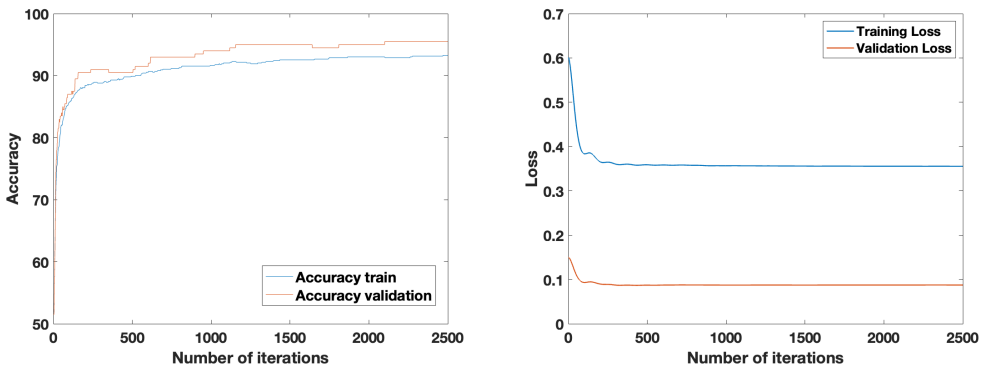


FIGURE 3. Training and validation accuracy and loss curves over iterations, showing convergence behavior of the model.

The confusion matrix of the proposed method, as reported in Table 6, is shown in Figure 4.

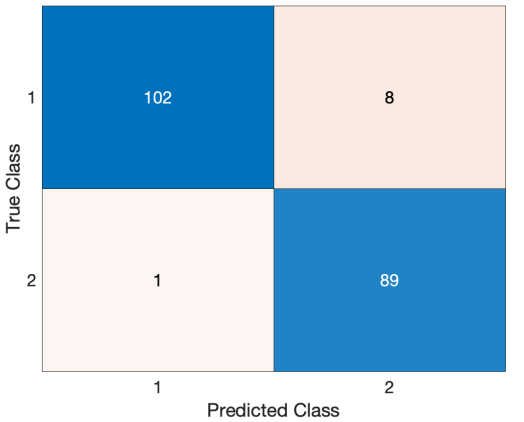


FIGURE 4. Confusion matrix showing the classification performance of the proposed method.

From Figure 4, it shows that the proposed method correctly classifies 102 samples as true negatives and 89 samples as true positives, indicating strong discriminative capability. The number of misclassifications is relatively small, with only 8 false positives and 1 false negative. This implies that the model maintains a high true positive rate while keeping the false negative rate very low, reflecting a stable classification structure. In particular, the minimal number of false negatives suggests that the model effectively captures the underlying decision boundary, whereas the limited false positives indicate that the model does not significantly overestimate the positive class.

Next we investigate the contribution of each feature to the classification performance by systematically removing one feature at a time and evaluating the resulting model using our proposed algorithm. Specifically, for each feature, we construct a reduced dataset by excluding that feature and then apply the proposed optimization-based learning framework to train the classifier. The performance is assessed in terms of accuracy (training and testing), precision, recall, F1-score, number of iterations, and computational time. This feature ablation strategy allows us to analyze the sensitivity of the model to individual features and to identify the most influential variables in the classification process. Formally, the stopping rule can be written as: terminate the algorithm at iteration k if $k \geq 600$ or if $\text{Acc}_{\text{train}} > 90\%$ and $\text{Acc}_{\text{test}} > 90\%$. The corresponding results are summarized in Table 8.

TABLE 8. Performance comparison based on incremental feature inclusion

Feature	Iter	Acc_{train}	Acc_{test}	Precision	Recall	F1	Time
Diabetic	600	89.6766	90.0000	88.0734	93.2039	90.5660	2.7223
AlcoholLevel	600	89.9254	90.0000	88.0734	93.2039	90.5660	2.7771
HeartRate	600	88.4328	90.0000	86.7257	95.1456	90.7407	2.7825
BloodOxygenLevel	285	90.0498	92.0000	88.4956	97.0874	92.5926	1.3113
BodyTemperature	600	87.6866	90.0000	84.8739	98.0583	90.9910	2.7401
Weight	600	92.5373	89.5000	84.1667	98.0583	90.5830	2.9926
MRI_Delay	262	91.9154	90.5000	85.0000	99.0291	91.4798	1.1999
Prescription	600	88.9303	90.0000	85.4701	97.0874	90.9091	2.7292
Dosage in mg	600	83.9552	82.5000	80.9091	86.4078	83.5681	2.7440
Age	410	90.0498	92.0000	89.1892	96.1165	92.5234	2.0925
Education_Level	449	90.0498	91.0000	87.6106	96.1165	91.6667	2.0661
Dominant_Hand	446	90.0498	90.5000	86.8421	96.1165	91.2442	2.5823
Gender	444	90.0498	90.5000	86.8421	96.1165	91.2442	2.0881
Family_History	499	90.6716	90.5000	86.2069	97.0874	91.3242	2.2725
Smoking_Status	435	90.0498	90.5000	86.8421	96.1165	91.2442	1.9877
APOE_ε4	457	90.0498	90.5000	86.8421	96.1165	91.2442	2.0912
Physical_Activity	426	90.0498	90.5000	86.8421	96.1165	91.2442	1.9773
Depression_Status	424	90.0498	90.5000	86.2069	97.0874	91.3242	1.9504
Cognitive_Test_Scores	600	87.5622	90.5000	85.0000	99.0291	91.4798	2.7499
Medication_History	320	90.0498	92.5000	87.2881	100.0000	93.2127	1.5302
Nutrition_Diet	300	90.0498	91.0000	85.7143	99.0291	91.8919	1.3936
Sleep_Quality	379	90.0498	91.5000	86.4407	99.0291	92.3077	1.7702
Chronic_Health_Conditions	409	90.0498	91.0000	86.3248	98.0583	91.8182	1.9220

Table 8 reports the classification performance obtained under the incremental inclusion of features, providing insight into the contribution of each variable to the predictive model. It is observed that the model maintains a relatively stable performance, with testing accuracy around 90%, even when only a small subset of features is employed. This indicates that basic clinical indicators already capture essential information for distinguishing between classes. As additional features are incorporated, particularly clinically informative variables such as BloodOxygenLevel and Age, the model exhibits a noticeable improvement in predictive performance. The best result is achieved when Medication_History is included, yielding the highest testing accuracy (92.5%), perfect recall (100%), and the largest F1-score. This suggests that such features play a dominant role in enhancing the discriminative capability of the model, especially in reducing false negatives, which is critical in medical screening tasks. In contrast, the inclusion of certain features, such as Dosage in mg, leads to a degradation in performance, indicating that not all variables contribute positively to the model and some may introduce redundancy or noise. Furthermore, beyond a certain level of feature inclusion, the improvement becomes marginal, implying that the model approaches a saturation point where additional features provide limited new information.

Next, we performed a feature analysis on the 23 selected features using the Classification Learner application in MATLAB. In particular, the feature selection process was conducted based on a feature ranking algorithm to evaluate the relative importance of each feature in the classification framework. The corresponding ranking scores obtained from the analysis are summarized in Table 9.

TABLE 9. Feature Ranking Algorithm

Rank	MRMR		Chi2		ReliefF		ANOVA	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score
1	Prescriptions	0.6927	Prescriptions	440.7479	Prescriptions	0.4267	Prescriptions	658.5953
2	Smoking_St	0.4890	DosageInMg	437.9007	Cognitive_Test	0.3014	DosageInMg	623.0436
3	Cognitive_Test	0.4885	Cognitive_Test	384.3445	DosageInMg	0.2332	Cognitive_Test	358.6872
4	DosageInMg	0.3129	Depression_St	175.2649	Depression_St	0.0986	Depression_St	213.9618
5	Depression_St	0.2823	APOE_4	94.0320	–	–	APOE_4	103.4314
6	APOE_4	0.2521	Smoking_St	11.5965	–	–	Smoking_St	78.2851
7	Family_His	0.1795	Education	8.9650	–	–	Education	10.3128

Based on the results presented in Table 9, the selected features were employed for the data classification process using our algorithm. The stopping criterion of the iterative procedure was determined by either reaching 2500 iterations or attaining $\mathrm{Acc}_{\mathrm{Train}} = 100$ and $\mathrm{Acc}_{\mathrm{Test}} = 100$. To investigate the effectiveness and stability of the proposed approach, the analysis was divided into the following cases.
Case 1 (C1) was designed to investigate the classification performance under different combinations of the highest-ranked features selected by the MRMR feature selection method. The considered feature sets are described as follows:

- C1F7: Prescriptions, Smoking_Status, Cognitive_Test_Scores, DosageInMg, Depression_Status, APO E_ε4 and Family_History
- C1F6: Prescriptions, Smoking_Status, Cognitive_Test_Scores, DosageInMg, Depression_Status and APOE_ε4
- C1F5: Prescriptions, Smoking_Status, Cognitive_Test_Scores, DosageInMg and Depression_Status
- C1F4: Prescriptions, Smoking_Status, Cognitive_Test_Scores and DosageInMg

Case 2 (C2) was conducted using feature subsets selected from the Chi2, ReliefF, and ANOVA feature selection methods in order to further examine the classification performance of the proposed algorithm under different feature ranking strategies as follows:

- C2F7: Prescriptions, DosageInMg, Cognitive_Test_Scores, Depression_Status, APOE_ε4 , Smoking_Status and Education_Level
- C2F6: Prescriptions, DosageInMg, Cognitive_Test_Scores, Depression_Status, APOE_ε4 and Smoking_Status
- C2F5: Prescriptions, DosageInMg, Cognitive_Test_Scores, Depression_Status and APOE_ε4
- C2F4: Prescriptions, DosageInMg, Cognitive_Test_Scores and Depression_Status
- C2F3: Prescriptions, DosageInMg and Cognitive_Test_Scores

The computational performance and classification results obtained from Case 1 and Case 2 are summarized in Table 10, while their corresponding visual comparisons are illustrated in Figures 5 and 6, respectively.

TABLE 10. Comparison of Iteration and CPU Time based on selected features

Cases	Measure	Features				
		7	**6**	**5**	**4**	**3**
C1	Iter	316	445	244	2500	–
	CPU	0.4081	0.6003	0.3430	3.1154	–
					$\mathrm{Acc}_{\mathrm{Train}} = 99.1294, \mathrm{Acc}_{\mathrm{Test}} = 99.5000$	
					$\mathrm{Pre} = 99.0385, \mathrm{Rec} = 100, \mathrm{F}_1 = 99.5169$	
C2	Iter	409	445	332	428	210
	CPU	0.5533	0.5646	0.4254	0.5860	0.2727

From Table 10, it can be observed that reducing the number of features can improve the computational performance of the proposed algorithm. In Case 1, the selection of five features yielded the best performance, achieving $\mathrm{Acc}_{\mathrm{Train}} = 100$ and $\mathrm{Acc}_{\mathrm{Test}} = 100$ with a computational time of 0.3430 seconds and requiring only 244 iterations. This result outperformed the cases employing fewer than five features as well as those using more than five features. Similarly, in Case 2, the selection of three features provided the most effective performance, attaining $\mathrm{Acc}_{\mathrm{Train}} = 100$ and $\mathrm{Acc}_{\mathrm{Test}} = 100$ with a computational time of 0.2727 seconds and only 210 iterations. It is noteworthy that both Case 1 and Case 2 consistently identified the same significant features, namely Prescriptions, DosageInMg, and Cognitive_Test_Scores. These features reflect the essential clinical and cognitive information associated with dementia screening, indicating their strong influence on the classification performance of the proposed model.

Based on the results presented in Table 10, the cases C1F5 and C2F5 were selected for further analysis. In particular, the training and validation accuracies were plotted to illustrate the learning behavior and convergence performance of the proposed algorithm. Furthermore, confusion matrices were constructed to demonstrate the classification performance and predictive capability of each considered case, as illustrated in Figures 5 and 6, respectively.

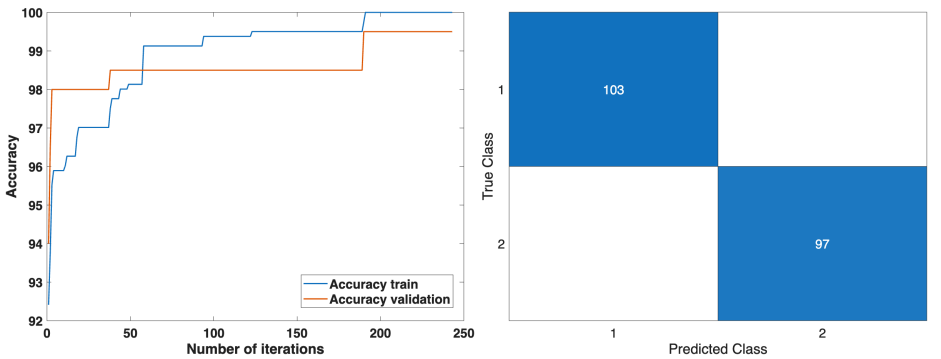


FIGURE 5. Training and validation accuracy and Confusion matrix showing the classification performance of the Case C1F5.

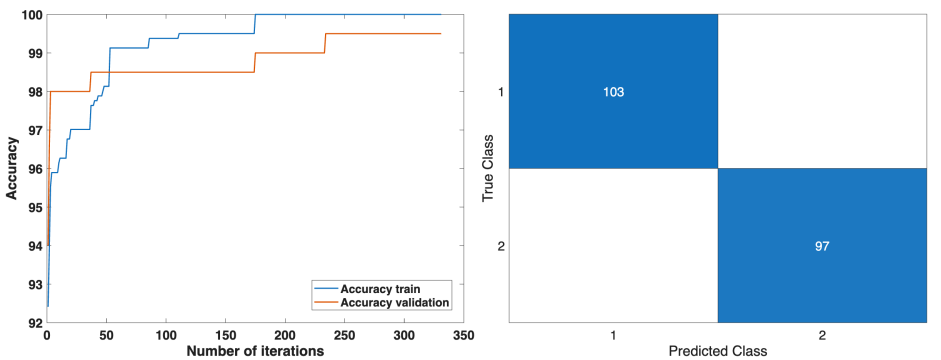


FIGURE 6. Training and validation accuracy and Confusion matrix showing the classification performance of the Case C2F5.

4. Conclusion

In this paper, we introduced a novel self-adaptive forward–backward algorithm for solving convex minimization problems in Hilbert spaces and investigated its convergence properties. By incorporating a self-adaptive mechanism for updating the control parameters, the proposed algorithm enhances the stability and efficiency of the iterative process. Under suitable conditions, both weak and strong convergence of the generated sequence were established. Numerical experiments in infinite-dimensional Hilbert spaces confirm the theoretical findings. The proposed method demonstrates robust convergence behavior for various initial points and achieves progressively smaller error levels, particularly for larger values of the parameter K, indicating improved approximation accuracy. Moreover, the practical performance of the algorithm was validated through its application to a dementia classification problem. When combined with the Extreme Learning Machine, the proposed method achieved superior classification performance compared to existing methods, including those of Combettes and Tseng. It yielded higher accuracy and F1-score, along with lower loss and computational time, highlighting its effectiveness in real-world scenarios.

Acknowledgments

Kunrada Kankam would like to express her sincere gratitude to Suan Dusit University for financial support under the Fundamental Fund (2025), grant no. FF69-3-213823. B. Tan thanks the supports of the Natural Science Foundation of Chongqing (No. CSTB2024NSCQ-MSX0354), the Fundamental Research Funds for the Central Universities (No. SWU-KQ24052 and No. SWU-KR25013), and the National Natural Science Foundation of China (No. 12471473).

Conflict of interest

The authors declare that they have no competing interests.

Data availability statement

Data Availability Statement The data supporting the findings of this study are publicly available from the Dementia Patient Health Dataset on Kaggle (https://www.kaggle.com/datasets/ timothyadeyemi/dementia-patient-health-dataset).

Institutional Review Board Statement

Ethical approval for this study was obtained from the Ethical Review Subcommittee for Human Research in Health Science, Research and Development Institute, Suan Dusit University (Approval No. SDU-RDI-HS 2026-019; Protocol No. HS025/2569). The study was approved on April 23, 2026, in accordance with the Declaration of Helsinki and ICH-GCP guidelines.

References

[1] Burachik, R. S.; Iusem, A. N. Enlargements of monotone operators. In *Set-Valued Mappings and Enlargements of Monotone Operators*, 161–220, Springer US, Boston, MA, 2008.

[2] Cholamjiak, P.; Van Hieu, D.; Cho, Y. J. Relaxed forward-backward splitting methods for solving variational inclusions and applications. *J. Sci. Comput.* **88** (2021), no. 3, Article 85.

[3] Cholamjiak, W.; Cholamjiak, P.; Suantai, S. An inertial forward-backward splitting method for solving inclusion problems in Hilbert spaces. *J. Fixed Point Theory Appl.* **20** (2018), no. 1, Article 42.

[4] Cholamjiak, W.; Yambangwai, D.; Das, S. Inertial three-step forward-backward splitting algorithms to solve inclusion problems and their application to image restoration problems. In *Mathematical Analysis for Engineering and Applied Sciences*, 141–172, CRC Press, 2025.

[5] Combettes, P. L.; Wajs, V. R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4** (2005), no. 4, 1168–1200.
[6] Combettes, P. L.; Pesquet, J. C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 185–212, Springer, New York, 2011.
[7] Costarelli, D.; Sambucini, A. R. A comparison among a fuzzy algorithm for image rescaling with other methods of digital image processing. *Constr. Math. Anal.* **7** (2024), no. 2, 45–68.
[8] Halpern, B. Fixed points of nonexpanding maps. *Bull. Amer. Math. Soc.* **73** (1967), no. 6, 957–961.
[9] Hanjing, A.; Suantai, S. Novel algorithms with inertial techniques for solving constrained convex minimization problems and applications to image inpainting. *Mathematics* **11** (2023), no. 8, Article 1813.
[10] He, S.; Yang, C. Solving the variational inequality problem defined on intersection of finite level sets. *Abstr. Appl. Anal.* **2013** (2013), no. 1, Article 942315.
[11] Huang, G. B.; Zhu, Q. Y.; Siew, C. K. Extreme learning machine: theory and applications. *Neurocomputing* **70** (2006), 489–501.
[12] Kankam, K.; Cholamjiak, W.; Cholamjiak, P. New inertial forward-backward algorithm for convex minimization with applications. *Demonstr. Math.* **56** (2023), no. 1, Article 20220188.
[13] Kankam, K.; Noor, M. A.; Cholamjiak, P. Three-step projected forward-backward algorithms for constrained minimization problem. *J. Appl. Math. Comput.* **71** (2025), no. 1, 465–487.
[14] Kankam, K.; Cholamjiak, P. A current forward-backward-forward method for inclusion problems. *Discrete Contin. Dyn. Syst. Ser. S* **18** (2025), no. 9, 2379–2396.
[15] Mokaba, L.; Abass, H. A.; Oyewole, O. K.; Malebana, K. P. Accelerated self-adaptive method for solving nonsmooth convex minimization problem in real Hilbert spaces. *Int. J. Anal. Appl.* **23** (2025), 160–160.
[16] Nakajo, K.; Takahashi, W. Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups. *J. Math. Anal. Appl.* **279** (2003), no. 2, 372–379.
[17] Opial, Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73** (1967), no. 4, 591–597.
[18] Suantai, S.; Noor, M. A.; Kankam, K.; Cholamjiak, P. Novel forward–backward algorithms for optimization and applications to compressive sensing and image inpainting. *Adv. Differ. Equ.* **2021** (2021), no. 1, Article 265.
[19] Suantai, S.; Thongsri, P. A fast forward-backward algorithm using linesearch and inertial techniques for convex bi-level optimization problems with applications. *Carpathian J. Math.* **40** (2024), no. 3, 769–788.
[20] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58** (1996), no. 1, 267–288.
[21] Tseng, P. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38** (2000), no. 2, 431–446.
[22] Yamada, I.; Ogura, N. Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings. *Numer. Funct. Anal. Optim.* **26** (2005), no. 7–8, 767–784.

[1] Elementary Education Program, Faculty of Education, Suan Dusit University Lampang Center, Lampang 52100, Thailand
Email address: kunrada_kan@dusit.ac.th

[2] Office of Research Administration, Chiang Mai University, Chiang Mai 50200, Thailand
Email address: suparat.ke@gmail.com

[3] Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
Email address: suparat.ke@gmail.com

[4] School of Mathematics and Statistics, Southwest University, Chongqing, China
Email address: bingtan@swu.edu.cn

[5] School of Science, University of Phayao, Phayao 56000, Thailand
Email address: prasit.ch@up.ac.th, watcharaporn.ch@up.ac.th

[6] MedUP Intelligence Healthcare and Data Innovation Center, School of Medicine, University of Phayao, Phayao 56000, Thailand
Email address: prasit.ch@up.ac.th, watcharaporn.ch@up.ac.th,pasa.su@up.ac.th

[7] School of Medicine, University of Phayao 56000, Thailand
Email address: pasa.su@up.ac.th