

Research Article

Breast Cancer Screening Using a Modified Inertial Projective Algorithms for Split Feasibility Problems

Pennipat Nabheerong,¹ Warissara Kiththiworaphongkich,² and Watcharaporn Cholamjiak,³

¹Radiology Department, School of Medicine, University of Phayao, Phayao 56000, Thailand ²Radiology Department, Phayao Hospital, Phayao 56000, Thailand ³School of Science, University of Phayao, Phayao 56000, Thailand

Correspondence should be addressed to Watcharaporn Cholamjiak; watcharaporn.ch@up.ac.th

Received 17 May 2023; Revised 30 July 2023; Accepted 17 August 2023; Published 7 September 2023

Academic Editor: Arjun Singh

Copyright © 2023 Pennipat Nabheerong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To detect breast cancer in mammography screening practice, we modify the inertial relaxed CQ algorithm with Mann's iteration for solving split feasibility problems in real Hilbert spaces to apply in an extreme learning machine as an optimizer. Weak convergence of the proposed algorithm is proved under certain mild conditions. Moreover, we present the advantage of our algorithm by comparing it with existing machine learning methods. The highest performance value of 85.03% accuracy, 82.56% precision, 87.65% recall, and 85.03% F1-score show that our algorithm performs better than the other machine learning models.

1. Introduction and Preliminaries

Breast cancer is common in women, with approximately 2 million diagnosed women worldwide each year [1]. Its mortality rate has increased over decades due to the change in risk factors, aging society, and better cancer registration and detection [2]. Many trials conclude that mammography screening at age 40s has mortality reduction, many years of life saved, and improved treatment, including evaluation of the extent of the disease [3]. However, there are some risks of this modality: overdiagnosis, false-positives, anxiety, and radiation injury [4]. Recently, imaging options combined with artificial intelligence are believed to be enhanced by integrating new screening protocols directed toward more personalized and precision medicine [1].

Many optimization algorithms were used to solve medical classification in machine learning; see [5, 6]. In this paper, we focus on the split feasibility problem (SFP) applying to mammography classification. Let \mathscr{C} and \mathscr{Q} be two nonempty closed and convex subsets of real Hilbert space such that $\mathscr{C} \subseteq \mathscr{H}_1$ and $\mathscr{Q} \subseteq \mathscr{H}_2$, and let $\mathscr{A} : \mathscr{H}_1 \longrightarrow \mathscr{H}_2$ be a bounded linear operator. The problem SFP is to

find
$$\omega^* \in \mathscr{C}$$
 such that $\mathscr{A}\omega^* \in \mathscr{Q}$, (1)

if such ω^* exist. The solution set $\Omega \coloneqq \{\omega^* \in \mathscr{C} : \mathscr{A}\omega^* \in \mathscr{Q}\}$ of the problem SFP (1) is denoted by Ω .

The first algorithm to solve the problem SFP (1) was presented by Censor and Elfving [7]. After that, many mathematicians (see [8–10]) applied the problem SFP (1) to solve many real-world problems such as machine learning, signal processing, image restoration, and many more. To find a solution of the problem SFP (1), Xu [11] proved that the point ω^* is a solution of the problem SFP (1) if and only if the point ω^* is a fixed point of the following mapping:

$$\operatorname{Proj}_{\mathscr{C}}(I - \lambda \nabla f) = \operatorname{Proj}_{\mathscr{C}}(I - \lambda \mathscr{A}^*(I - \operatorname{Proj}_{\mathscr{Q}})\mathscr{A}).$$
(2)

Later on, Byrne [12] was the first to propose a popular \mathscr{CQ} algorithm for solving the problem SFP (1). The \mathscr{CQ} algorithm of Byrne [12] was generated as follows:

$$\omega_{n+1} = \operatorname{Proj}_{\mathscr{C}}(\omega_n - \lambda \mathscr{A}^* (I - \operatorname{Proj}_{\mathscr{Q}}) \mathscr{A} \omega_n), \forall n \ge 1, \quad (3)$$

where the parameter λ belongs in the interval $(0, 2/ \|\mathcal{A}\|^2)$ which makes the mapping $(I - \lambda \mathcal{A}^* (I - \operatorname{Proj}_{\widehat{\mathcal{Q}}}) \mathcal{A}$ is nonexpansive where *I* is an identity, i.e.,

$$\begin{aligned} & \| (I - \lambda \mathscr{A}^* (I - \operatorname{Proj}_{\widehat{\alpha}}) \mathscr{A} x - (I - \lambda \mathscr{A}^* (I - \operatorname{Proj}_{\widehat{\alpha}}) \mathscr{A} y \| \\ & \leq \| x - y \|, \forall x, y \in \mathscr{H}_1, \end{aligned}$$

$$(4)$$

and \mathscr{A}^* denotes for the adjoint operator of \mathscr{A} , while $\operatorname{Proj}_{\mathscr{C}}$ and $\operatorname{Proj}_{\widehat{\mathscr{Q}}}$ are the orthogonal projections onto \mathscr{C} and $\widehat{\mathscr{Q}}$, respectively. The overall cost of calculation is not high if the metric projections onto \mathscr{C} and $\widehat{\mathscr{Q}}$ are simple to calculate. However, precisely computing the metric projection is difficult or requires too much effort in some circumstances when \mathscr{C} and $\widehat{\mathscr{Q}}$ are complex constructs.

Later on, Yang [13] introduced a relaxed \mathscr{CQ} algorithm by modifying the \mathscr{C} and \mathscr{Q} sets of the \mathscr{CQ} algorithm (3) to reveal sets as follows:

$$\mathscr{C}_{n} = \{ x \in \mathscr{H}_{1} : c(\omega_{n}) \leq \langle \xi_{n}, \omega_{n} - x \rangle \} \text{ and } \mathscr{Q}_{n}$$

= $\{ y \in \mathscr{H}_{2} : q(\mathscr{A}\omega_{n}) \leq \langle \eta_{n}, \mathscr{A}\omega_{n} - y \rangle \},$ (5)

where $c: \mathscr{H}_1 \longrightarrow \mathbb{R} \cup \{+\infty\}$ and $q: \mathscr{H}_2 \longrightarrow \mathbb{R} \cup \{+\infty\}$ are two proper convex functions such that $\xi_n \in \partial c(\omega_n)$ and $\eta_n \in \partial q(\mathscr{A}\omega_n)$. Since the projections $\operatorname{Proj}_{\mathscr{G}_n}$ and $\operatorname{Proj}_{\mathscr{Q}_n}$ are easier to use, many mathematicians use them to modify numerous algorithms for solving the problem SFP (1); see [14–16].

One of the techniques to speed up the convergence of the algorithms is the inertial technique which Polyak first introduced [17] in 1964. Polyak's algorithm was called the heavy ball method, and it was improved by Nesterov [18]. Later on, it has been widely used to solve a wide variety of problems in the optimization field, as seen in [9, 19–22].

In 2017, Dang et al. [9] modified the inertial technique which was introduced by Alvarez and Attouch [23] with the CQ algorithm of the Byrne algorithm (3) for the problem SFP (1) in a real Hilbert space. This algorithm was defined as follows:

$$\rho_n = \omega_n + \sigma_n (\omega_n - \omega_{n-1}), \tag{6}$$

$$\omega_{n+1} = \operatorname{Proj}_{\mathscr{C}_n} \left(\rho_n - \lambda \mathscr{A}^* \left(I - \operatorname{Proj}_{\widehat{\mathscr{Q}}_n} \right) \mathscr{A}(\rho_n) \right), \forall n \ge 1,$$
(7)

where the parameter λ is in the interval involving the norm of operator \mathscr{A} , \mathscr{C}_n and \mathscr{Q}_n are the revel set introduced by Yang [13], and the extrapolation factor $\sigma_n \in [0, \bar{\sigma}_n]$ and $\sigma \in [0, 1)$ such that

$$\bar{\sigma}_{n} = \min\left\{\sigma, \frac{1}{\max\left\{n^{2} \|\omega_{n} - \omega_{n-1}\|^{2}, n^{2} \|\omega_{n} - \omega_{n-1}\|\right\}}\right\}, \forall n \ge 1.$$
(8)

The weak convergence of algorithm $\{\omega_n\}$ generated by (6) was proved under the conditions of the extrapolation factor (8) and the stepsize parameter λ .

Very recently, Wang and Yu [32] generalized an inertial relaxed \mathscr{CQ} of Yang [13] by modifying \mathscr{C}_n , \mathscr{Q}_n as follows:

$$\mathscr{C}_{n} = \left\{ x \in \mathscr{H}_{1} : c(\omega_{n}) \leq \langle \xi_{n}, \omega_{n} - x \rangle - \frac{\alpha}{2} \|\omega_{n} - x\|^{2} \right\},$$
$$\mathscr{Q}_{n} = \left\{ y \in \mathscr{H}_{2} : q(\mathscr{A}\omega_{n}) \leq \langle \eta_{n}, \mathscr{A}\omega_{n} - y \rangle - \frac{\beta}{2} \|\mathscr{A}\omega_{n} - y\|^{2} \right\}.$$
(9)

 \mathscr{C}_n and \mathscr{Q}_n of Wang and Yu [24] can be reduced to \mathscr{C}_n and \mathscr{Q}_n of Yang [13] when α and β are set to 0. The inertial generalized relaxed \mathscr{CQ} algorithm (IGRCQ) of Wang and Yu [24] was introduced by $\omega_0, \omega_1 \in \mathscr{H}_1$, and

$$\rho_n = \omega_n + \sigma_n (\omega_n - \omega_{n-1}), \tag{10}$$

$$\omega_{n+1} = \operatorname{Proj}_{\mathscr{C}_n}(\rho_n - \lambda_n \nabla f_n(\rho_n)), \forall n \ge 1,$$
(11)

where $\{\sigma_n\} \in [0,\sigma) \in [0,1), \, \{\varepsilon_n\} \in (0,4)$ and

$$\lambda_{n} = \begin{cases} \varepsilon_{n} \frac{f_{n}(\omega_{n})}{\left\|\nabla f_{n}(\omega_{n})\right\|^{2}}, & \text{if } \left\|\nabla f_{n}(\omega_{n})\right\| \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$
(12)

For each $n \in \mathbb{N}$, the functions are defined as follows:

$$f_{n}(\cdot) = \frac{1}{2} \left\| \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathscr{A}(\cdot) \right\|^{2},$$
(13)

$$\nabla f_n = \mathscr{A}^* \left(I - \operatorname{Proj}_{\mathcal{Q}_n} \right) \mathscr{A}.$$
(14)

It was shown that, under the conditions $\sum_{n=1}^{\infty} \sigma_n \|\omega_n - \omega_{n-1}\|^2 < \infty$ and $\liminf_{n \to \infty} \varepsilon_n (4 - \varepsilon_n) > 0$, the sequence $\{\omega_n\}$ created by (10) weakly converges to a solution of the problem SFP (1).

In this paper, we modify the inertial technique with relaxed \mathscr{CQ} algorithms and Mann's algorithm to solve the split feasibility problems in Hilbert spaces. We establish weak convergence theorems under suitable conditions. We apply our main result to solve a data classification problem by using extreme learning machine with mammographic mass data set from UCI, and then compare the performance of our algorithm with other existing machine learning algorithms.

2. Main Results

In this section, we introduce a new modified inertial projective algorithm by combining Mann algorithms with inertial technique and inertial relaxed \mathscr{CQ} algorithm. Assume that \mathscr{C} and \mathscr{Q} are two nonempty closed and convex subsets of real Hilbert space such that $\mathscr{C} \subseteq \mathscr{H}_1$ and $\mathscr{Q} \subseteq \mathscr{H}_2$ such that

$$\mathcal{C} = \{ \omega \in \mathcal{H}_1 : c(\omega) \le 0 \}, \, \mathcal{Q} = \{ \omega \in \mathcal{H}_2 : q(\mathcal{A}\omega) \le 0 \}, \ (15)$$

where $\mathscr{A}: \mathscr{H}_1 \longrightarrow \mathscr{H}_2$ is a bounded linear operator, $c: \mathscr{H}_1 \longrightarrow \mathbb{R}$ and $q: \mathscr{H}_2 \longrightarrow \mathbb{R}$ are lower semicontinuous convex functions. We next assume that ∂c and ∂q are bounded operators. For a sequence $\{y_n\}$ in \mathscr{H}_1 , we modify the half-spaces \mathscr{C}_n and $\widehat{\mathcal{Q}}_n$ by using the idea of Wang and Yu [24] as follows:

$$\mathscr{C}_{n} = \left\{ \omega \in \mathscr{H}_{1} : c(y_{n}) \leq \langle \varrho_{n}, y_{n} - \omega \rangle - \frac{\alpha}{2} \left\| y_{n} - \omega \right\|^{2} \right\}, \quad (16)$$

where $\varrho_n \in \partial c(y_n)$, $\alpha \ge 0$, and

$$\mathcal{Q}_{n} = \left\{ \omega \in \mathcal{H}_{2} : q(\mathcal{A}y_{n}) \leq \langle \xi_{n}, \mathcal{A}y_{n} - \omega \rangle - \frac{\beta}{2} \| \mathcal{A}y_{n} - \omega \|^{2} \right\},$$
(17)

where $\xi_n \in \partial q(\mathscr{A}y_n)$ and $\beta \ge 0$. We see that $\mathscr{C} \subseteq \mathscr{C}_n$ and $\mathscr{Q} \subseteq \mathscr{Q}_n$ for each $n \ge 1$. Define $f_n(\cdot)$ and $\nabla f_n(\cdot)$ as in (13).

We now introduce a modified inertial projective algorithms as follows:

Assume that the following condition hold:

(C1) $\sum_{n=1}^{\infty} \sigma_n \max \{ \|\omega_n - \omega_{n-1}\|^2, \|\omega_n - \omega_{n-1}\| \} < \infty;$ (C2) $0 < \liminf_{n \to \infty} \lambda_n \le \limsup_{n \to \infty} \lambda_n < 2/\|\mathscr{A}\|^2;$ (C3) $0 < \liminf_{n \to \infty} \alpha_n \le \limsup_{n \to \infty} \alpha_n < 1.$

Theorem 1. Let \mathcal{H}_1 and \mathcal{H}_2 be two real Hilbert spaces, and let \mathcal{C} and \mathcal{Q} be nonempty closed convex subsets such that $\mathcal{C} \subseteq \mathcal{H}_1$ and $\mathcal{Q} \subseteq \mathcal{H}_2$. Let $\mathcal{A} : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ be a bounded linear operator. Assume that the solution set Ω of the problem SFP (1) is nonempty, the condition (C1)-(C2) hold. Then the sequence $\{\omega_n\}$ generated by Algorithm 1:. converges weakly to a point $\omega^* \in \Omega$.

Proof. Let $\omega^* \in \Omega$. Since $(I - Proj_{\hat{\omega}_n})$ is firmly nonexpansive, then for each $n \in \mathbb{N}$, we have

$$2f_{n}(y_{n}) = \left\| \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}y_{n} \right\|^{2}$$

$$= \left\| \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}y_{n} - \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}\omega^{*} \right\|^{2}$$

$$\leq \left\langle \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}y_{n} - \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}\omega^{*}, \mathcal{A}y_{n} - \mathcal{A}\omega^{*} \right\rangle$$

$$= \left\langle \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}y_{n}, \mathcal{A}y_{n} - \mathcal{A}\omega^{*} \right\rangle$$

$$= \left\langle \mathcal{A}^{*} \left(I - \operatorname{Proj}_{\mathcal{Q}_{n}} \right) \mathcal{A}y_{n}, y_{n} - \omega^{*} \right\rangle$$

$$= \left\langle \nabla f_{n}(y_{n}), y_{n} - \omega^{*} \right\rangle.$$
(18)

In the other hand, we set $t_n = y_n - \lambda_n \nabla f_n(y_n)$, From (18), we have

$$\begin{split} \|\omega_{n+1} - \omega^{*}\|^{2} &= \|(1 - \alpha_{n})y_{n} + \alpha_{n}z_{n} - \omega^{*}\|^{2} \\ &\leq (1 - \alpha_{n})\|y_{n} - \omega^{*}\|^{2} + \alpha_{n}\|z_{n} - \omega^{*}\|^{2} \\ &\leq (1 - \alpha_{n})\|y_{n} - \omega^{*}\|^{2} \\ &+ \alpha_{n} \left(\|t_{n} - \omega^{*}\|^{2} - \|t_{n} - z_{n}\|^{2}\right) \\ &= \|y_{n} - \omega^{*}\|^{2} - \alpha_{n}\|y_{n} - z_{n}\|^{2} \\ &+ 2\alpha_{n}\lambda_{n} \langle \nabla f_{n}(y_{n}), y_{n} - z_{n} \rangle - 4\alpha_{n}\lambda_{n}f_{n}(y_{n}) \\ &\leq \|y_{n} - \omega^{*}\|^{2} - \alpha_{n}\|y_{n} - z_{n}\|^{2} \\ &+ 2\alpha_{n}\lambda_{n}\|\nabla f_{n}(y_{n})\|\|y_{n} - z_{n}\| - 4\alpha_{n}\lambda_{n}f_{n}(y_{n}) \\ &\leq \|y_{n} - \omega^{*}\|^{2} + 2\lambda_{n}^{2}\alpha_{n}\|\mathscr{A}\|^{2}f_{n}(y_{n}) \\ &- 4\alpha_{n}\lambda_{n}f_{n}(y_{n}) = \|y_{n} - \omega^{*}\|^{2} \\ &- 4\lambda_{n}\alpha_{n}\left(1 - \frac{1}{2}\lambda_{n}\|\mathscr{A}\|^{2}\right)f_{n}(y_{n}), \end{split}$$
(19)

$$\begin{aligned} \|y_{n} - \omega^{*}\|^{2} &= \|\rho_{n} - \lambda_{n} \nabla f_{n}(\rho_{n}) - \omega^{*}\|^{2} \\ &\leq \|\rho_{n} - \omega^{*}\|^{2} + \lambda_{n}^{2} \|\nabla f_{n}(\rho_{n})\|^{2} \\ &- 2\lambda_{n} \langle \rho_{n} - \omega^{*}, \nabla f_{n}(\rho_{n}) - \nabla f_{n}(\omega^{*}) \rangle \\ &\leq \|\rho_{n} - \omega^{*}\|^{2} + \lambda_{n}^{2} \|\nabla f_{n}(\rho_{n})\|^{2} - \frac{2\lambda_{n}}{\|\mathscr{A}\|^{2}} \|\nabla f_{n}(\rho_{n})\|^{2} \\ &\leq \|\rho_{n} - \omega^{*}\|^{2} - \left(\frac{2\lambda_{n}}{\|\mathscr{A}\|^{2}} - \lambda_{n}^{2}\right) \|\nabla f_{n}(\rho_{n})\|^{2} \\ &\leq \|\omega_{n} - \omega^{*}\|^{2} + 2\sigma_{n} \langle \omega_{n} - \omega_{n-1}, y_{n} - \omega^{*} \rangle \\ &- \left(\frac{2\lambda_{n}}{\|\mathscr{A}\|^{2}} - \lambda_{n}^{2}\right) \|\nabla f_{n}(\rho_{n})\|^{2}. \end{aligned}$$

$$(20)$$

Replacing (20) into (19), we have

$$\begin{split} \|\omega_{n+1} - \omega^*\|^2 &\leq \|\omega_n - \omega^*\|^2 + 2\sigma_n \langle \omega_n - \omega_{n-1}, y_n - \omega^* \rangle \\ &- \left(\frac{2\lambda_n}{\|\mathscr{A}\|^2} - \lambda_n^2\right) \|\nabla f_n(\rho_n)\|^2 \\ &- 4\lambda_n \alpha_n \left(1 - \frac{1}{2}\lambda_n \|\mathscr{A}\|^2\right) f_n(y_n). \end{split}$$

$$\tag{21}$$

This implies that

$$\begin{pmatrix} \frac{2\lambda_n}{\|\mathscr{A}\|^2} - \lambda_n^2 \end{pmatrix} \|\nabla f_n(\rho_n)\|^2 + 4\lambda_n \alpha_n \left(1 - \frac{1}{2}\lambda_n \|\mathscr{A}\|^2\right) f_n(y_n)$$

$$\leq \|\omega_n - \omega^*\|^2 - \|\omega_{n+1} - \omega^*\|^2 + 2\sigma_n \langle \omega_n - \omega_{n-1}, y_n - \omega^* \rangle.$$
(22)

```
Initialization: Take \omega_0, \omega_1 \in \mathcal{C} and set n = 1.

Iterative Steps: Calculate \{\omega_n\} by the following step:

Step 1. Compute

\rho_n = \omega_n + \sigma_n(\omega_n - \omega_{n-1}),

where 0 \le \sigma_n < \sigma < 1.

Step 2. Compute

y_n = \rho_n - \lambda_n \nabla f_n(\rho_n), z_n = Proj_{\mathcal{C}_n}(y_n - \lambda_n \nabla f_n(y_n)),

where \lambda_n \in (0, (2/||\mathcal{A}||^2)).

Step 3. Compute

\omega_{n+1} = (1 - \alpha_n)y_n + \alpha_n z_n,

where \alpha_n \in (0, 1).

Replace n with n + 1 and then repeat Step 1.
```

ALGORITHM 1: Inertial modified relaxed CQ Mann algorithm (IMRCQM).

Since $I - \lambda_n \nabla f_n$ and $\operatorname{Proj}_{\mathscr{C}_n}$ are nonexpansive, we have

$$\begin{aligned} \|\omega_{n+1} - \omega^*\| &= \|(1 - \alpha_n)y_n + \alpha_n z_n - \omega^*\| \\ &\leq (1 - \alpha_n)\|y_n - \omega^*\| + \alpha_n\|z_n - \omega^*\| \\ &\leq \|\rho_n - \omega^*\| \leq \|\omega_n - \omega^*\| + \sigma_n\|\omega_n - \omega_{n-1}\|. \end{aligned}$$
(23)

By Lemma 1 in [25] and (C1), we have that $\{\|\omega_n - \omega^*\|\}$ is convergence sequence for any $\omega^* \in \Omega$. Therefore, $\{\omega_n\}$ is bounded. From the definition of $\{\rho_n\}$, $\{\rho_n\}$ is also bounded. It follows from (20), (C1)-(C3) that

$$\lim_{n \longrightarrow \infty} \|\nabla f_n(\rho_n)\| = \lim_{n \longrightarrow \infty} f_n(y_n) = 0.$$
 (24)

Again by $I - \lambda_n \nabla f_n$ and $P_{\mathscr{C}_n}$ are nonexpansive, then we have

$$\begin{split} \|\omega_{n+1} - \omega^*\|^2 &= \|(1 - \alpha_n)y_n + \alpha_n z_n - \omega^*\|^2 \\ &\leq \|\rho_n - \omega^*\|^2 - (1 - \alpha_n)\alpha_n\|y_n - z_n\|^2 \\ &\leq \|\omega_n - \omega^*\|^2 + 2\sigma_n \langle \omega_n - \omega_{n-1}, \rho_n - \omega^* \rangle \\ &- (1 - \alpha_n)\alpha_n\|y_n - z_n\|^2, \end{split}$$
(25)

which implies that

$$(1 - \alpha_n)\alpha_n \|y_n - z_n\|^2 + 2\sigma_n \langle \omega_{n-1} - \omega_n, \rho_n - \omega^* \rangle \leq \|\omega_n - \omega^*\|^2 - \|\omega_{n+1} - \omega^*\|^2.$$
(26)

From $\lim_{n \to \infty} \|\omega_n - \omega^*\|$ exists, (C1) and (C3), we obtain

$$\lim_{n \to \infty} \|y_n - z_n\| = 0.$$
 (27)

It follows from (C2) and (24) that

$$\lim_{n \to \infty} \|y_n - \rho_n\| = \lim_{n \to \infty} \lambda_n \|\nabla f_n(\rho_n)\| = 0.$$
 (28)

And it is clearly from (C1), we have

$$\lim_{n \to \infty} \|\rho_n - \omega_n\| = \lim_{n \to \infty} \sigma_n \|\omega_n - \omega_{n-1}\| = 0.$$
 (29)

From (28) and (29), we obtain

$$\lim_{n \to \infty} \|y_n - \omega_n\| = 0.$$
(30)

Finally, let ω^* be a weak sequential cluster point of $\{\omega_n\}$. There exists a subsequence $\{\omega_{n_k}\}$ of $\{\omega_n\}$ which converges weakly to $\omega^* \in \mathcal{H}_1$. From (30), we also have that $\{y_{n_k}\}$ converges weakly to ω^* and hence $\mathcal{A}y_{n_k} \rightharpoonup \mathcal{A}\omega^*$ as $k \longrightarrow \infty$. By the definition of z_n , we have that $P_{\hat{Q}_n}(Ay_n) \in \hat{Q}_n$. This implies that

$$q\left(\mathscr{A}y_{n_{k}}\right) \leq \left\langle \xi_{n_{k}}, \left(I - \operatorname{Proj}_{\mathscr{Q}_{n_{k}}}\right) \mathscr{A}a_{n_{k}} \right\rangle - \frac{\beta}{2} \left\| \left(I - \operatorname{Proj}_{\mathscr{Q}_{n_{k}}}\right) \mathscr{A}a_{n_{k}} \right\|^{2},$$

$$(31)$$

where $\xi_{n_k} \in \partial q(\mathscr{A}y_{n_k})$. By our assumption as ∂q is bounded (24) and (31), we have $q(\mathscr{A}\omega^*) \leq 0$, this shows that $\mathscr{A}\omega^* \in \mathscr{Q}$. Again, by the definition of $\{z_n\}$, we have that $z_{n_k} \in \mathscr{C}_{n_k}$. This implies that

$$c\left(y_{n_{k}}\right) \leq \left\langle \varrho_{n_{k}}, y_{n_{k}} - z_{n_{k}} \right\rangle - \frac{\alpha}{2} \left\| y_{n_{k}} - z_{n_{k}} \right\|^{2}$$
(32)

where $Q_{n_k} \in \partial c(y_{n_k})$. By our assumption as ∂c is bounded (27) and (32), we have $c(\omega^*) \leq 0$, this shows that $\omega^* \in \mathscr{C}$. By Opial's lemma in [26], we can conclude that $\{\omega_n\}$ converges weakly to a solution in Ω . This completes the proof.

3. Application to Data Classification Problem

Nowadays, many cancer patients are reported around the world each year. In the population survey in Global Cancer

Attribute	Туре	\overline{x}	Standard deviation	Max	Min	Coefficient of variation
BI-RADS	Double	4.33	0.63	6	1	14.65
Age	Double	55.78	14.67	96	18	26.30
Shape	Double	2.78	1.24	4	1	44.66
Margin	Double	2.81	1.57	5	1	55.71
Density	Double	2.92	0.35	4	1	12.04
Severity	Category					

TABLE 1: Overview of mammographic mass data set from UCI.

Statistics 2020, information was found that breast cancer was the most severe disease with 258 new cases per hour [27]. Breast cancer is more common in developed countries than in developing countries, and the number of cases varies with per capita income (GDP per capita). In addition, people's way of life in the city (urbanization) and those environments result in more risk behaviors for breast cancer. Importantly, it also found that the number of patients tends to increase significantly each year as well. In Thailand, breast cancer is the 1st most common cancer among females and the 3rd most common among both males and females. There are 8,266 deaths from breast cancer per year or about 1 person per hour. The above data shows that breast cancer is rapidly increasing and directly threatening the female population globally, including Thailand. In addition, the number of doctors specializing in such diseases is limited, not enough to provide services to the patients. Therefore, the use of technology-based knowledge related to artificial intelligence or machine learning including deep learning is the basis for creating tools or innovations that are efficient and accurate in assisting medical personnel in screening and diagnosing breast cancer. As a result, patients will be screened quickly and accurately, cured in the early stages and could reduce mortality. It is also an indicator of the modernization of the country's development in the field of public health in the future. Mammography plays a central part in the early detection of breast cancers because it can show changes in the breast years before a patient or physician can feel them. Research has shown that annual mammograms lead to early detection of breast cancers when they are most curable, and breast-conservation therapies are available. Women, beginning at age 40, should screen for mammography every year. In this research, we use the mammographic mass dataset from UCI is available on the UCI website. (https://archive.ics.uci.edu/ml/datasets/ Mammographic+Mass?fbclid=IwAR1TL44iSKmqXX6PMiS VjqGVZRD-suQTPEVsejq01SUylZwildNu7UWEPZQ). This dataset contains a BI-RADS assessment, the patient's age, and three BI-RADS attributes: shape of mass, margin of mass, and density of mass together with the ground truth (the severity field) for 516 benign and 445 malignant masses that have been identified on full-field digital mammogram collected at the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006. These datasets can indicate how well a several computer-aided diagnosis (CAD) system performs compared to the radiologists [28]. After 167 missing attribute values from 7 BI-RADS, 5 ages, 31 shapes, 48 margins, and 76 densities were removed before

the training process. The following Table 1 shows the overview of all attributes.

In 2021, Parvez et al. [29] showed many machine learning predictive models to classify breast cancer using this mammographic mass dataset. The following Table 2 shows the comparison of our algorithm 1 consider in two constrain closed convex sets L_1 and L_2 with machine learning predictive models by Parvez et al. [29] after feature engineering.

From Table 2, the results show that our algorithm 1 when constrain closed convex set by L_2 was used gives the highest accuracy 85.03% after removing rows with missing values and outliers. We next explain how our algorithm 1 optimizes weight parameter in training data for machine learning. We focus on extreme learning machine (ELM) by using 5-fold cross-validation [30]. The ELM method is defined as follows: assume that $\mathcal{U} \coloneqq \{(\mu_s, r_s): \mu_s \in \mathbb{R}^n, r_s \in \mathbb{R}^m, s = 1, 2, \dots, N\}$ is a set of training data and r_s is a target. Finding optimal output weight using the output function is the objective of the ELM method. The following output function is for single-hidden layer feed-forward neural networks (SLFNs) with M hidden nodes:

$$O_s = \sum_{i=1}^{M} w_i \mathcal{V}(\langle c_i, \mu_s \rangle + e_i), \qquad (33)$$

where \mathcal{V} is an activation function and c_i and e_i are parameters of weight and finally the bias, respectively. The optimal output weight w_i at the *i*-th hidden node is found by setting the hidden layer output matrix \mathcal{H} as follows:

$$\mathscr{H} = \begin{bmatrix} \mathscr{V}(\langle c_1, \mu_1 \rangle + e_1) & \cdots & \mathscr{V}(\langle c_M, \mu_1 \rangle + e_M) \\ \vdots & \ddots & \vdots \\ \mathscr{V}(\langle c_1, \mu_N \rangle + e_1) & \cdots & \mathscr{V}(\langle c_M, \mu_N \rangle + e_M) \end{bmatrix}.$$
(34)

We also assume an optimal output weight $w = [w_1^T, \dots, w_M^T]^T$ such that $\mathscr{H}w = \mathscr{R}$, where $\mathscr{R} = [r_1^T, \dots, r_N^T]^T$ is the training target data. For solving linear system $\mathscr{H}w = \mathscr{R}$, we use the least square problem when the *Moore-Penrose generalized inverse* of \mathscr{H} is not easy to find. To ovoid overfitting in the machine learning, we consider constrain least square

Machine learning model	Original dataset (%)	Highest accuracy (%)	Data cleaning or feature engineering		
Logistic regression	80.46	82.50	Removing two least contributed features, RFE, or removing two least contributed features, correlation matrix		
Linear discriminant analysis	78.90	84.50	Removing two least contributed features, RFE, or removing two least contributed features, correlation msatrix		
K-nearest neighbors	79.03	82.05	Removing two least contributed features, RFE, or removing two least contributed features, correlation matrix		
Classification and regression trees	74.48	82.97	Removing two least contributed features, RFE, or removing two least contributed features, correlation matrix		
Gaussian Naive Bayes	78.37	83.73	Removing rows with missing values and outliers		
Support vector machines	80.34	83.73	Removing two least contributed features, RFE, or removing two least contributed features, correlation matrix		
Algorithm 1 (L_1)	53.89	83.23	Removing rows with missing values and outliers		
Algorithm 1 (L_2)	53.89	85.03	Removing rows with missing values and outliers		

TABLE 2: Highest accuracy of ML algorithms after feature engineering.

TABLE 3: All different necessary parameters of each algorithm.

Parameter	σ	λ_n	α_n	γ	λ	ε_n
Algorithm 1 (L_1)	0.9999	$\frac{0.9999}{\max \ \operatorname{eig}(A^T A)))}$	$\frac{1}{1.2}$	7	_	
Algorithm 1 (L_2)	0.9999	0.9999/max $\operatorname{eig}(A^{T}A)))$	$\frac{1}{1.2}$	17	_	_
Algorithm (6) (L_1)	0.9999	_	_	7	0.9999/max $\operatorname{eig}(A^T A)))$	_
Algorithm (6) (L_2)	0.9999	_	_	17	0.9999/max $\operatorname{eig}(A^T A)))$	_
Algorithm (10) (L_1)	0.9999	_	_	7	_	0.1
Algorithm (10) (L_2)	0.9999	_	_	17	—	0.1

TABLE 4: All performances of each algorithm for comparison.

Algorithm	Number of iterations	Training time	Accuracy	Precision	Recall	F1-score
Algorithm (6) (L_1)	410	0.1614	82.04	81.40	83.33	82.35
Algorithm (6) (L_2)	271	0.1439	74.25	94.19	68.07	79.02
Algorithm (10) (L_1)	188	0.2072	82.63	80.23	85.19	82.63
Algorithm (10) (L_2)	242	0.3689	77.25	62.79	90.00	73.97
Algorithm 1 (L_1)	81	0.0694	83.23	80.23	86.25	83.13
Algorithm 1 (L_2)	259	0.2027	85.03	82.56	87.65	85.03

problem in two different closed convex subsets of ${\mathscr H}$ as follows:

$$\min_{\boldsymbol{\omega}\in\mathscr{C}_{1}} \{ \|\mathscr{H}\boldsymbol{\omega} - \mathscr{R}\|_{2}^{2} \},
\min_{\boldsymbol{\omega}\in\mathscr{C}_{2}} \{ \|\mathscr{H}\boldsymbol{\omega} - \mathscr{R}\|_{2}^{2} \},$$
(35)

where $\mathscr{C}_1 = \{x \in \mathscr{H} : ||x||_1 \le \gamma\}, \mathscr{C}_2 = \{x \in \mathscr{H} : ||x||_2^2 \le \gamma\}$ such that γ is regularization parameters. Setting $f(\omega) = 1/2$ $||(I - Proj_{\mathscr{Q}})\mathscr{H}\omega||_2^2, Q = Q_1 = Q_2 = \{\mathscr{R}\}, c_1(\omega) = ||\omega||_1 - \gamma, c_2$ $(\omega) = ||\omega||_2^2 - \gamma$, and $q(\omega) = 1/2 ||\omega - \mathscr{R}||^2$ for our algorithm 1 to solve the problem (3.1). We use four evaluation metrics: accuracy, precision, recall, and F1-score [31] as explained below for comparing the performance of the classification algorithms.

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

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

$$Recall = \frac{TP}{TN + FN} \times 100\%.$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall},$$
(36)



FIGURE 1: Training and validation loss and the accuracy plots of Algorithm 1: with constrain closed convex set L1.



FIGURE 2: Training and validation loss and the accuracy plots of Algorithm 1: with constrain closed convex set L_2 .

where TN is the true negative, FP is the false positive, FN is the false negative, and TP is the true positive.

For avoiding model overfitting, we consider accuracy and loss plots. This research, we use the following binary cross-entropy loss function:

$$\operatorname{Loss} = -\frac{1}{m} \sum_{i=1}^{m} O_i \log \widehat{y}_i + (1 - O_i) \log \left(1 - \widehat{O}_i\right), \qquad (37)$$

where \widehat{O}_i is the *i*-th scalar value in the model output, O_i is the corresponding target value, and *m* is the number of scalar values in the model output.

For comparison with other existing methods from the literature, the necessary parameters of each algorithm are chosen in Table 3. The extrapolation parameter of algorithm 1 (L_1 , L_2) and algorithm (10) is in the following from:

$$\sigma_{n} = \begin{cases} \frac{\sigma}{n^{2} \max\left\{\left\|\omega_{n} - \omega_{n-1}\right\|^{2}, \left\|\omega_{n} - \omega_{n-1}\right\|\right\}}, & \text{if } n > N, \omega_{n} \neq \omega_{n-1}, \\ \sigma & \text{otherwise,} \end{cases}$$
(38)

where *N* is a number of iterations that we want to stop and $\sigma_n = \bar{\sigma}_n$ for algorithm (6).

Sigmoid is set as an activation function with hidden nodes M = 160, and four evaluation metrics of each algorithm are shown in Table 4.

Table 4 shows that our algorithm 1 with constrain closed convex set L_2 is the highest F1-score, precision, recall, and accuracy efficiency. Additionally, our algorithm 1 with constrain closed convex set L_1 has the lowest number of iterations. The optimal-fitting of our algorithm 1 is shown by considering the training and validation loss with the accuracy.

From Figures 1 and 2, we observe that both of algorithm 1 L_1 and L_2 have optimal-fitting models. This means that the algorithm suitably learns the training dataset and generalizes well to predict the severity of mammographic mass based on BI-RADS assessment, the patient's age, shape, margin, and density of mass.

Remark 2. Since a matrix \mathcal{H} in ELM was generated by a finite dataset that contains real numbers, thus we can see from Table 4 that our Algorithm 1:. which requires a norm estimation of the bounded linear operator of \mathcal{H} , gives more efficiency than the algorithm (10) of Wang and Yu [24].

4. Conclusion and Discussion

Nowadays, there are many studies interested in the accuracy of artificial intelligence (AI) for the detection of breast cancer in mammography screening programme. Some believe that artificial intelligence (AI) has helped improve radiologists' performance and provides results equivalent or superior to those of radiologists' alone such as reduce the volume in screen-reading without affecting cancer detection substantially [32]. Although there are some issues that should be more explore including possible factors on recall and interval cancers [33], Freeman et al. [34] performed the systematic review of test accuracy and concluded that there is inadequate evidence in judgement of accuracy of artificial intelligence (AI) in detecting breast cancer on screening mammography. There is still small researches, which could not be representing the real effect of artificial intelligence (AI) in clinical practice or where on the clinical pathway AI might be of most benefit.

This paper presented an applying inertial modified relaxed *CQ* Mann algorithms for split feasibility problems for extreme learning machine based on BI-RADS assessment, the patient's age, and three BI-RADS attributes for predicting the severity of mammographic mass lesion to assist the physician regarding making decision about whether to go for biopsy or not. The comparison with other machine learning models and existing algorithms for split feasibility problems shows that our algorithm provides the highest performance value of 85.03% accuracy, 82.56% precision, 87.65% recall, and 85.03% F1-score. Moreover, considering training and validation loss, and the accuracy plots show that our algorithm has good fit model.

Data Availability

The mammographic mass dataset from UCI is available on the UCI website (http://archive.ics.uci.edu/ml/datasets/ mammographic+mass).

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

The writing, which includes review and editing, was done by P.N.. Review and editing were in charge of W.K.. Writing, which includes the original draft and software, was the responsible of W.C.. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was supported by the National Research Council of Thailand and University of Phayao (N42A650334) and Thailand Science Research and Innovation, University of Phayao (FF66-UoE).

References

- R. M. Mann, R. Hooley, R. G. Barr, and L. Moy, "Novel approaches to screening for breast cancer," *Radiology*, vol. 297, no. 2, pp. 266–285, 2020.
- [2] S. Lukasiewicz, M. Czeczelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanislawek, "Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review," *Cancers*, vol. 13, no. 17, p. 4287, 2021.
- [3] G. Durhan, A. Azizova, Ö. Önder et al., "Imaging findings and clinicopathological correlation of breast cancer in women under 40 years old," *European Journal of Breast Health*, vol. 15, no. 3, pp. 147–152, 2019.
- [4] L. J. Grimm, C. S. Avery, E. Hendrick, and J. A. Baker, "Benefits and risks of mammography screening in women ages 40 to 49 years," *Journal of Primary Care & Community Health*, vol. 13, article 215013272110583, 2022.
- [5] P. Peeyada, H. Dutta, K. Shiangjen, and W. Cholamjiak, "A modified forward-backward splitting methods for the sum of two monotone operators with applications to breast cancer prediction," *Mathematical Methods in the Applied Sciences*, vol. 46, no. 1, pp. 1251–1265, 2023.
- [6] P. Sarnmeta, W. Inthakon, D. Chumpungam, and S. Suantai, "On convergence and complexity analysis of an accelerated forward-backward algorithm with linesearch technique for convex minimization problems and applications to data prediction and classification," *Journal of Inequalities and Applications*, vol. 2021, no. 1, pp. 1–20, 2021.
- [7] Y. Censor and T. Elfving, "A multiprojection algorithm using Bregman projections in a product space," *Numerical Algorithms*, vol. 8, no. 2, pp. 221–239, 1994.
- [8] Q. H. Ansari and A. Rehan, "Split feasibility and fixed point problems," in *Nonlinear Analysis*, pp. 281–322, Birkhäuser, New Delhi, 2014.
- [9] Y. Dang, J. Sun, H. Xu, School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China, and Department of Mathematics and Statistics, Curtin University, Perth, WA 6102, Australia, "Inertial accelerated algorithms for solving a split feasibility problem," *Journal of Industrial & Management Optimization*, vol. 13, no. 3, pp. 1383–1394, 2017.
- [10] S. Suantai, S. Kesornprom, and P. Cholamjiak, "A new hybrid & algorithm for the split feasibility problem in Hilbert spaces and its applications to compressed sensing," *Mathematics*, vol. 7, no. 9, p. 789, 2019.
- [11] H. K. Xu, "Iterative methods for the split feasibility problem in infinite-dimensional Hilbert spaces," *Inverse Problems*, vol. 26, no. 10, article 105018, 2010.
- [12] C. Byrne, "Iterative oblique projection onto convex sets and the split feasibility problem," *Inverse Problems*, vol. 18, no. 2, pp. 441–453, 2002.
- [13] Q. Yang, "The relaxed CQ algorithm solving the split feasibility problem," *Inverse Problems*, vol. 20, no. 4, pp. 1261–1266, 2004.
- [14] A. Gibali, D. T. Mai, and N. T. Vinh, "A new relaxed CQ algorithm for solving split feasibility problems in Hilbert spaces and its applications," *Journal of Industrial & Management Optimization*, vol. 15, no. 2, pp. 963–984, 2019.
- [15] S. Kesornprom, N. Pholasa, and P. Cholamjiak, "On the convergence analysis of the gradient-CQ algorithms for the split

feasibility problem," *Numerical Algorithms*, vol. 84, no. 3, pp. 997–1017, 2020.

- [16] S. Suantai, P. Peeyada, A. Fulga, and W. Cholamjiak, "Heart disease detection using inertial Mann relaxed CQ algorithms for split feasibility problems," *AIMS Mathematics*, vol. 8, no. 8, pp. 18898–18918, 2023.
- [17] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," USSR Computational Mathematics and Mathematical Physics, vol. 4, no. 5, pp. 1–17, 1964.
- [18] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," *Doklady Akademii Nauk SSSR*, vol. 269, pp. 543–547, 1983.
- [19] Q. L. Dong, J. Z. Huang, X. H. Li, Y. J. Cho, and T. M. Rassias, "MiKM: multi-step inertial Krasnosel'skii–Mann algorithm and its applications," *Journal of Global Optimization*, vol. 73, no. 4, pp. 801–824, 2019.
- [20] P. E. Maingé, "Inertial iterative process for fixed points of certain quasi-nonexpansive mappings," *Set-Valued Analysis*, vol. 15, no. 1, pp. 67–79, 2007.
- [21] Y. Shehu and A. Gibali, "New inertial relaxed method for solving split feasibilities," *Optimization Letters*, vol. 15, no. 6, pp. 2109–2126, 2021.
- [22] S. Suantai, N. Pholasa, and P. Cholamjiak, "Relaxed & algorithms involving the inertial technique for multiple-sets split feasibility problems," *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, vol. 113, no. 2, pp. 1081–1099, 2019.
- [23] F. Alvarez and H. Attouch, "An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping," *Set-Valued Analysis*, vol. 9, no. 1/2, pp. 3–11, 2001.
- [24] F. Wang and H. Yu, "An inertial relaxed CQ algorithm with an application to the LASSO and elastic net," *Optimization*, vol. 70, no. 5-6, pp. 1101–1119, 2021.
- [25] K. K. Tan and H. K. Xu, "Approximating fixed points of nonexpansive mappings by the Ishikawa iteration process," *Journal of Mathematical Analysis and Applications*, vol. 178, no. 2, pp. 301–308, 1993.
- [26] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, 2011.
- [27] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [28] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process," *Medical Physics*, vol. 34, no. 11, pp. 4164–4172, 2007.
- [29] M. Parvez, R. B. Eshun, M. Bikdash, and A. K. Islam, "A Machine Learning Predictive Model to Classify Severity of Breast Cancer Based on Mammographic Mass Dataset," 2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2021, pp. 81–87, Laguna Hills, CA, USA, 2021.
- [30] V. A. Kumari and R. Chitra, "Classification of diabetes disease using support vector machine," *International Journal of Engineering Research and Applications*, vol. 3, no. 2, pp. 1797– 1801, 2013.

- [31] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques, Waltham, MA*, vol. 10, Morgan Kaufman Publishers, 2012.
- [32] Y. W. Chang, J. K. An, N. Choi et al., "Artificial intelligence for breast cancer screening in mammography (AI-STREAM): a prospective multicenter study Design in Korea Using AIbased CADe/x," *Journal of Breast Cancer*, vol. 25, no. 1, pp. 57–68, 2022.
- [33] M. Larsen, C. F. Aglen, S. R. Hoff, H. Lund-Hanssen, and S. Hofvind, "Possible strategies for use of artificial intelligence in screen-reading of mammograms, based on retrospective data from 122,969 screening examinations," *European Radiology*, vol. 32, no. 12, pp. 8238–8246, 2022.
- [34] K. Freeman, J. Geppert, C. Stinton et al., "Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy," *BMJ*, vol. 374, 2021.