Results in Nonlinear Analysis 6 (2023) No. 3, 1–18
https://doi.org/10.31838/rna/2023.06.03.001
Available online at www.nonlinear-analysis.com

Results in Nonlinear Analysis
ISSN 2636-7556

Peer Reviewed Scientific Journal

A modified subgradient extragradient method for equilibrium problems to predict prospective mathematics teachers' digital proficiency level

Raweerote Suparatulatorn[a], Watcharaporn Cholamjiak[b], Nipa Jun-on[c]

[a]Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand; [b]School of Science, University of Phayao, Phayao 56000, Thailand; [c]Department of Mathematics, Faculty of Science, Lampang Rajabhat University, Lampang 52100, Thailand.

Abstract

Numerous research in the field of education analytics has attempted to discover a significant indicator and predictor of the digital proficiency level of pre-service teachers. While university course alterations in their academic performance are perceived as ordinary, significant fluctuations in their academic performance in courses related to digital technology may require further investigation and consideration, particularly regarding their digital proficiency level. However, such a method is problematic due to the complexities of describing digital academic paths. In this paper, we modify the extragradient method with an inertial extrapolation step and viscosity-type method to solve equilibrium problems of the pseudomonotone bifunction operator. Under the assumption that the bifunction satisfies the Lipchitz-type condition in real Hilbert spaces, we obtain strong convergence theorem. Next, we apply our algorithm to classify the digital proficiency level of pre-service teachers in order to investigate the correlation between academic achievement in digital technology-related courses and digital proficiency level. Finally, we establish several situations in which the digital proficiency level of pre-service teachers might either increase or decrease.

Keywords: Strong convergence, Pseudomonotone equilibrium problem, Data classification problem, Digital proficiency level, Educational data classification

AMS Subject Classification: 65K15, 47H05, 49M37, 97M10, 97C70, 97U70

Email addresses: raweerote.s@gmail.com (Raweerote Suparatulatorna); c-wchp007@hotmail.com (Watcharaporn Cholamjiak); nipa.676@g.lpru.ac.th (Nipa Jun-on)*

Received May 9, 2023; Accepted June 24, 2023; Online July 20, 2023

1. Introduction

Digital technologies are regarded not only as instruments for working and studying but also as tools for social involvement. Mathematics teachers must build their own and their students' digital abilities due to the pervasiveness of digital technology in our society and prevent the continued widening of the digital divide [1]. At the same time, the vulnerability of primary education in mathematics has been identified concerning the alarmingly low technical competency at lower educational levels [2]. In the COVID-19 pandemic period, the complicated demands and current deficiencies in this sector have become readily apparent.

The COVID-19 pandemic touched all levels and aspects of education, including teacher preparation and professional development [3, 4]. To cope with emergency distance learning and teaching, both teacher educators and pre-service teachers had to overcome difficult technological, pedagogical, social, cognitive, and practical hurdles [5, 6]. Teacher preparation institutions were obliged to construct new infrastructure and learning environments and adapt to new teaching techniques, learning situations, and study materials. Moreover, pre-service teachers as prospective teachers were required to strengthen their digital skills in order to meet all the new obstacles.

Particularly, to fully realize the benefits of digital technology in the mathematics classroom [7], prospective mathematics teachers need substantial preparation and training. Consequently, the most recent prospective mathematics teacher preparation requirements involve incorporating digital technology into mathematics instruction to generate competent potential mathematics teachers [8]. Using legitimate classification technique to evaluate the digital proficiency level of prospective mathematics teachers offers a practical foundation for eligibility into the profession [9, 10].

This paper studies the *equilibrium problem* (EP), initially introduced by Muu and Oettli [11]. The problem EP is to find an element z^* in a nonempty closed convex subset C of a real Hilbert space $\mathcal{H}$ such that

$$f(z^*, y) \geq 0, \quad \forall y \in C, \tag{1.1}$$

where $f : C \times C \to \mathbb{R}$ be a bifunction with $f(x,x) = 0$ for all $x \in \mathcal{H}$, and $EP(C,f)$ is denoted for a solution set of EP (1.1).

Inspired by the modified algorithm to solve variational inequalities of Korpelevich [12] which is called *extragradient method* (EM), Tran et al. [13] proposed the *two-step extragradient method* (TSEM) for solving EP (1.1). The weak convergence was proved under the standard condition; the control stepsize μ needs to belong in the control interval, making the bifunction f Lipschitz. However, in the process of the two-step, extragradient method [13], projection onto nonempty closed convex set C was used in two steps for each iteration. It isn't straightforward and can affect the method's efficiency if C has a complex structure. This method is defined $x_0 \in \mathcal{H}$ and $x_1 \in C$ and

$$\begin{aligned} y_n &= \arg\min_{y \in C} \left\{ \mu f(x_n, y) + \frac{1}{2} \| x_n - y \|^2 \right\}, \\ x_{n+1} &= \arg\min_{y \in C} \left\{ \mu f(y_n, y) + \frac{1}{2} \| x_n - y \|^2 \right\}. \end{aligned} \tag{1.2}$$

A natural question that arises in the case of infinite dimensional Hilbert spaces is how to design an algorithm that uses easier and provides strong convergence. This question was made clearly by Hieu [14] with *Halpern subgradient extragradient method* (HSEM), which was modified from the HSEM of Kraikaew and Saejung [15] for variational inequalities. This method is defined by $u \in \mathcal{H}$ and

$$\begin{aligned} \rho_n &= \arg\min_{y \in C} \left\{ \mu f(x_n, y) + \frac{1}{2} \| x_n - y \|^2 \right\}, \\ \mathcal{H}_n &= \{ v \in \mathcal{H} : \langle (x_n - \mu t_n) - \rho_n, v - \rho_n \rangle \}, \quad t_n \in \partial_2 f(x_n, \rho_n), \\ y_n &= \arg\min_{y \in \mathcal{H}_n} \left\{ \mu f(\rho_n, y) + \frac{1}{2} \| x_n - y \|^2 \right\}, \\ x_{n+1} &= \eta_n u + (1 - \eta_n) y_n, \end{aligned} \tag{1.3}$$

where μ is still some constant depending on the interval that makes the bifunction *f* satisfies the Lipschitz condition and $\{\eta_n\} \subset (0,1)$ which satisfies the principle conditions

$$\lim_{n\to\infty} \eta_n = 0, \quad \sum_{n=1}^{\infty} \eta_n = \infty. \tag{1.4}$$

Another famous method that makes the algorithm converge strongly is the viscosity approximation method which was introduced by Moudafi [16]. Using the Moudafi [16]'s idea, Muangchoo [17] combined a viscosity-type method with the extragradient algorithm for obtaining strong convergence theorem of the EP (1.1) such that *f* is pseudomonotone. This method is defined by

$$\begin{aligned} \rho_n &= \arg\min_{y\in C} \left\{ \mu_n f(x_n, y) + \frac{1}{2}\|x_n - y\|^2 \right\}, \\ \mathcal{H}_n &= \{v \in \mathcal{H} : \langle (x_n - \mu_n t_n) - \rho_n, v - \rho_n \rangle\}, \quad t_n \in \partial_2 f(x_n, \rho_n), \\ y_n &= \arg\min_{y\in \mathcal{H}_n} \left\{ \sigma\mu_n f(\rho_n, y) + \frac{1}{2}\|x_n - y\|^2 \right\}, \\ x_{n+1} &= \eta_n \nabla\phi(x_n) + (1-\eta_n) y_n, \end{aligned} \tag{1.5}$$

where $\sigma \in (0,\sigma) \subset \left(0, \min\left\{1, \frac{1}{2\ell_1}, \frac{1}{2\ell_2}\right\}\right)$, $\nabla\phi$ is a contraction function on $\mathcal{H}$ with constant $\rho \in [0,1)$, $\{\eta_n\}$ satisfies the principle conditions (1.4), and the update stepsize $\{\mu_n\}$ satisfies the following:

$$\mu_{n+1} = \begin{cases} \min\left\{\sigma, \dfrac{\sigma f(\rho_n, y_n)}{K_n}\right\}, & \text{if } \dfrac{\sigma f(\rho_n, y_n)}{K_n} > 0, \\ \lambda_1, & \text{otherwise}, \end{cases} \tag{1.6}$$

where $K_n = f(x_n, y_n) - f(x_n, \rho_n) - \ell_1 \|x_n - \rho_n\|^2 - \ell_2 \|y_n - \rho_n\|^2 + 1$ and ℓ_1, ℓ_2 are Lipschitz constants of *f*.

Another way to study the convergence of algorithms is to find a method that makes algorithms converge faster. Polyak [18] first introduced the inertial technique in 1964 for convergence speed up. This algorithm was generated for solving convex minimization. Shehu et al. [19] recently modified the inertial technique with the Halpern-type algorithm and subgradient extragradient method for obtaining strong convergence to a solution of $EP(f,C)$ such that *f* is pseudomonotone. The step-sizes $\{\mu_n\}$ is developed by updating the step-sizes method without knowing the Lipschitz-type constants of the bifunction *f*. This method is defined by $u \in \mathcal{H}$ and

$$\begin{aligned} w_n &= \eta_n u + (1-\eta_n) x_n + \delta_n (x_n - x_{n-1}), \\ \rho_n &= \arg\min_{y\in C} \left\{ \lambda_n f(w_n, y) + \frac{1}{2}\|w_n - y\|^2 \right\}, \\ \mathcal{H}_n &= \{v \in \mathcal{H} : \langle (w_n - \lambda_n t_n) - \rho_n, v - \rho_n \rangle\}, \quad t_n \in \partial_2 f(w_n, \rho_n), \\ y_n &= \arg\min_{y\in \mathcal{H}_n} \left\{ \lambda f(\rho_n, y) + \frac{1}{2}\|w_n - y\|^2 \right\}, \\ x_{n+1} &= \tau w_n + (1-\tau) y_n, \end{aligned} \tag{1.7}$$

where the inertial parameter $\delta_n \in \left[0, \frac{1}{3}\right)$, $\tau \in \left(0, \frac{1}{2}\right]$, the update stepsize $\{\mu_n\}$ satisfies the following:

$$\mu_{n+1} = \begin{cases} \min\left\{ \dfrac{\sigma}{2} \dfrac{\|w_n - \rho_n\|^2 + \|y_n - \rho_n\|^2}{f(w_n, y_n) - f(w_n, \rho_n) - f(\rho_n, y_n)}, \mu_n \right\}, & \text{if } f(w_n, y_n) - f(w_n, \rho_n) - f(\rho_n, y_n) > 0, \\ \mu_n, & \text{otherwise}, \end{cases} \tag{1.8}$$

$\sigma \in (0,1)$ and $\{\eta_n\}$ still satisfies the principle conditions (1.4). This update step-size $\{\mu_n\}$ is limited in the computation and can not be modified in another way.

Inspired by the previous works, we introduce a new modified subgradient extragradient method for obtaining strong convergence to a solution of $EP(f,C)$ using viscosity-type methods. In applications, we apply our algorithm to solve the classification problem of pre-service teachers' digital proficiency level using a dataset of 474 pre-service mathematics and sciences teachers responsible for mathematics classrooms in their school practicum from five cohorts. Pre-service teacher data using nine attributes, including major; gender; type of supplementary digital technology training courses (short, medium, and long format); grade point of Information Technology for Learning course; grade point of Innovation and Information Technology course; grade point of Digital Technology for teaching mathematics course; Grade Point Average (GPA); and digital proficiency level.

2. Preliminaries

2.1 Digital proficiency level

This study measured the digital proficiency level of participants based on DigCompEdu framework [20] that classify pre-service teachers into six levels as the following.

(*A*1) Newcomer: Newcomers know the potential of digital technology to enhance educational and professional activity. However, they have minimal experience with digital technologies and utilize them primarily for lesson planning, administration, and organizational communication. Newcomers require direction and support to extend their repertoire and use their present digital skills in the instructional sphere.

(*A*2) Explorer: Explorers are conscious of the possibilities of digital technologies and are eager to explore them in order to improve educational and professional practice. They have begun adopting digital technology in certain areas of digital competency without a consistent strategy. Explorers need encouragement, insight, and inspiration, such as via the direction and example of colleagues, as part of a cooperative sharing of practices.

(*B*1) Integrator: Integrators experiment with digital technology in many situations and for various goals, incorporating them into many of their activities. They use them inventively to increase multiple parts of their professional involvement. They are keen to extend their practicing repertoire. However, they still attempt to comprehend which tools perform best in specific contexts and to adopt digital technology to pedagogical ideas and practices. To become experts, Integrators need additional time for exploration and contemplation, along with collaborative support and information sharing.

(*B*2) Expert: Experts skillfully, creatively, and critically employ various digital technologies to better their professional tasks. They pick digital technology for specific scenarios and attempt to comprehend the advantages and disadvantages of different digital tactics. They are inquisitive and receptive to new ideas, understanding there are many items they have not yet tried. They use experimentation to broaden, organize, and consolidate their repertory of techniques. Regarding innovative practice, experts are the cornerstone of every educational organization.

(*C*1) Leader: Leaders use digital technology consistently and comprehensively to improve pedagogical and professional activities. They depend on an extensive repertory of digital tactics, from which they can choose the best applicable for every circumstance. They continually reflect on their procedures and strive to improve them. Through interaction with peers, they remain current on new advances and ideas. They serve as a source of motivation for those to whom they impart their knowledge.

(*C*2) Pioneer: Pioneers challenge the sufficiency of current digital and educational methods, of which they are Leaders. They are worried about the limitations or disadvantages of these techniques and are motivated to reinvent education even more. Pioneers explore extremely innovative and complicated digital technology or build new instructional strategies. They are distinct and uncommon types, innovators, and role models for younger educators.

In this paper, we left out the C2 level since it needs longitudinal experiences in a teaching career and is a unique and rare type of teacher. Thus, the class that we want to predict by our method will include only from (A1) Newcomer to (C1) Leader level.

2.2 Mathematical background

In what follows, recall that $\mathcal{H}$ is a real Hilbert space. Let C be a nonempty, closed and convex subset of $\mathcal{H}$. We denote $\rightharpoonup$ and $\to$ as weak and strong convergence, respectively. We next collect some necessary definitions and lemmas for proving our main results. For $u \in \mathcal{H}$, define the metric projection P_C from $\mathcal{H}$ onto C by

$$P_C u := arg \min_{v \in C} \|u - v\|.$$

It has been known for the fact that P_C can be distinguished by the inequality

$$\langle u - P_C u, v - P_C u \rangle \leq 0 \tag{2.1}$$

for any $u \in \mathcal{H}$ and $v \in C$. Next, the following equalities and inequality are valid for inner product spaces. Assume $u, v \in \mathcal{H}$,

$$\|u + v\|^2 \leq \|u\|^2 + 2\langle v, u + v \rangle, \tag{2.2}$$

$$\|au + (1 - a)v\|^2 = a\|u\|^2 + (1 - a)\|v\|^2 - a(1 - a)\|u - v\|^2 \tag{2.3}$$

for any $a \in \mathbb{R}$.

A normal cone of C at $x \in C$ is defined by

$$N_C(x) = \{z \in \mathcal{H} : \langle z, y - x \rangle \leq 0, \text{ forall } y \in C\}.$$

Let $g : C \to \mathbb{R}$ be a convex function and subdifferential of g at $x \in C$ is defined by

$$\partial g(x) = \{z \in \mathcal{H} : g(y) - g(x) \geq \langle z, y - x \rangle, \text{ for all } y \in C\}.$$

A bifunction $f : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ on C is stated to be

i. pseudomonotone if $f(u, v) \geq 0 \Rightarrow f(v, u) \leq 0$, for all $u, v \in C$;
ii. satisfies the Lipschitz-like criteria for some $\ell_1, \ell_2 > 0$, the following inequality is satisfied

$$f(u, w) \leq f(u, v) + f(v, w) + \ell_1 \|u - v\|^2 + \ell_2 \|v - w\|^2, \text{ for all } u, v, w \in C.$$

Lemma 2.1: [21] *Let $g : C \to \mathbb{R}$ be a subdifferentiable, convex and lower semi-continuous function on C. An element $x \in C$ is a minimizer of a function g if and only if*

$$0 \in \partial g(x) + N_C(x),$$

where $\partial g(x)$ stands for the subdifferential of g at $x \in C$ and $N_C(x)$ the normal cone of C at x.

Lemma 2.2: [22] *Let $\{a_n\}$ and $\{c_n\}$ be nonnegative sequences of real numbers such that $\sum_{n=1}^{\infty} c_n < \infty$, and let $\{b_n\}$ be a sequence of real numbers such that* $\limsup_{n\to\infty} b_n \leq 0$. *If for any $n \in \mathbb{N}$ such that*

$$a_{n+1} \leq (1 - \gamma_n)a_n + \gamma_n b_n + c_n,$$

where $\{\gamma_n\}$ is a sequence in (0, 1) *such that $\sum_{n=1}^{\infty} \gamma_n = \infty$, then* $\lim_{n\to\infty} a_n = 0$.

Lemma 2.3: [23] *Let $\{\Gamma_n\}$ be a sequence of real numbers such that there exists a subsequence $\{\Gamma_{n_j}\}_{j\in\mathbb{N}}$ of $\{\Gamma_n\}$ satisfying $\Gamma_{n_j} < \Gamma_{n_j+1}$ for all $j \in \mathbb{N}$. Define a sequence of integers $\{\psi(n)\}_{n\geq n^*}$ by*

$$\psi(n) := \max\{k \leq n : \Gamma_k < \Gamma_{k+1}\}. \tag{2.4}$$

Then $\{\psi(n)\}_{n\geq n^*}$ is a nondecreasing sequence such that $\lim_{n\to\infty}\psi(n) = \infty$, and for all $n \geq n^*$, we have that $\Gamma_{\psi(n)} \leq \Gamma_{\psi(n)+1}$ and $\Gamma_n \leq \Gamma_{\psi(n)+1}$.

3. Main result

To study the convergence analysis, consider the following conditions.

($\mathcal{C}1$) The solution set $EP(f,C)$ is nonempty and f is pseudomonotone on C;
($\mathcal{C}2$) f meet the Lipschitz-like condition on $\mathcal{H}$ through $\ell_1 > 0$ and $\ell_2 > 0$;
($\mathcal{C}3$) $f(z,\cdot)$ is subdifferentiable and convex on $\mathcal{H}$ for each fixed $z \in \mathcal{H}$;
($\mathcal{C}4$) $\limsup_{n\to\infty} f(z_n, y) \leq f(z^*, y)$ for each $y \in C$ and $\{z_n\} \subset C$ satisfies $z_n \rightharpoonup z^*$.

Algorithm 1

Initialization: Select arbitrary elements $x_0, x_1 \in \mathcal{H}$. Given $0 < \mu_n \leq \mu < \min\left\{\frac{1}{2\ell_1}, \frac{1}{2\ell_2}\right\}$. Let $\{\delta_n\} \subset [0,\infty)$ and $\{\eta_n\} \subset (0,1)$ satisfies

$$\lim_{n\to\infty}\eta_n = 0, \sum_{n=1}^{\infty}\eta_n = \infty \quad \textit{and} \quad \lim_{k\to\infty}\frac{\delta_n}{\eta_n}\|x_n - x_{n-1}\| = 0.$$

Iterative Steps: Construct $\{x_n\}$ by using the following steps:

Step 1. Set

$$\rho_n = \eta_n\nabla\phi(x_n) + (1-\eta_n)x_n + \delta_n(x_n - x_{n-1})$$

where $\phi : \mathcal{H} \to \mathbb{R}$ is a differentiable function such that $\nabla\phi$ is contraction with constant $\rho \in [0,1)$.

Step 2. Compute

$$y_n = \arg\min_{y\in C}\left\{\mu_n f(\rho_n, y) + \frac{1}{2}\|\rho_n - y\|^2\right\}.$$

If $\rho_n = y_n$, then stop. Otherwise

Step 3. Evaluate

$$x_{n+1} = \arg\min_{y\in\mathcal{H}_n}\left\{\mu_n f(y_n, y) + \frac{1}{2}\|\rho_n - y\|^2\right\},$$

where $w_n \in \partial_2 f(\rho_n, y_n)$ satisfying $\rho_n - \mu_n w_n - y_n \in N_C(y_n)$ and construct a half-space

$$\mathcal{H}_n = \{z \in \mathcal{H} : \langle \rho_n - \mu_n w_n - y_n, z - y_n\rangle \leq 0\}.$$

Replace n by $n+1$ and then repeat **Step 1.**

Lemma 3.1: *Let* $\rho_n = y_n$ *in Algorithm 1, then* $\rho_n \in EP(f,C)$.

Proof. By the definition of y_n with Lemma 2.1, we have

$$0 \in \partial_2\left(\mu_n f(\rho_n,\cdot) + \frac{1}{2}\|\rho_n - \cdot\|^2\right)(y_n) + N_C(y_n)$$

Thus, we can write $\mu_n\tilde{w}_n + y_n - \rho_n + \bar{w}_n = 0$, where $\tilde{w}_n \in \partial_2 f(\rho_n, y_n)$ and $\bar{w}_n \in N_C(y_n)$. Due to $\rho_n = y_n$ implies that $\mu_n\tilde{w}_n + \bar{w}_n = 0$. Thus, we have

$$\mu_n\langle\tilde{w}_n, y - y_n\rangle + \langle\bar{w}_n, y - y_n\rangle = 0$$

for all $y \in C$. By $\bar{w}_n \in N_C(y_n)$ implies $\langle\bar{w}_n, y - y_n\rangle \le 0$ for all $y \in C$ and through above expression, we obtain

$$\mu_n\langle\tilde{w}_n, y - y_n\rangle \ge 0 \tag{3.1}$$

for all $y \in C$. Due to $\tilde{w}_n \in \partial_2 f(\rho_n, y_n)$ and using the subdifferential definition, we obtain

$$\langle\tilde{w}_n, y - y_n\rangle \le f(\rho_n, y) - f(\rho_n, y_n) \tag{3.2}$$

for all $y \in C$. From the inequalities (3.1) and (3.2) with $0 < \mu_n \le \mu$ implies that $f(\rho_n, y) \ge 0$ for all $y \in C$, that is, $\rho_n \in EP(f,C)$.

Suppose that $f : \mathcal{H}\times\mathcal{H} \to \mathbb{R}$ meet the items $(\mathcal{C}1)-(\mathcal{C}3)$, we have

$$\|x_{n+1} - \varpi\|^2 + (1 - 2\ell_1\mu_n)\|\rho_n - y_n\|^2 + (1 - 2\ell_2\mu_n)\|y_n - x_{n+1}\|^2 \le \|\rho_n - \varpi\|^2 \tag{3.3}$$

for all $\varpi \in EP(f,C)$.

Proof. Let $\varpi \in EP(f,C)$, then by using Lemma 2.1, we have

$$0 \in \partial_2\left(\mu_n f(y_n,\cdot) + \frac{1}{2}\|\rho_n - \cdot\|^2\right)(x_{n+1}) + N_{\mathcal{H}_n}(x_{n+1})$$

Thus, we can write $\mu_n\tilde{w}_n + x_{n+1} - \rho_n + \bar{w}_n = 0$, where $\tilde{w}_n \in \partial_2 f(y_n, x_{n+1})$ and $\bar{w}_n \in N_{\mathcal{H}_n}(x_{n+1})$. This implies that

$$\langle\rho_n - x_{n+1}, y - x_{n+1}\rangle = \mu_n\langle\tilde{w}_n, y - x_{n+1}\rangle + \langle\bar{w}_n, y - x_{n+1}\rangle$$

for all $y \in \mathcal{H}_n$. Given that $\bar{w}_n \in N_{\mathcal{H}_n}(x_{n+1})$ then $\langle\bar{w}_n, y - x_{n+1}\rangle \le 0$ for all $y \in \mathcal{H}_n$. Therefore, we have

$$\langle\rho_n - x_{n+1}, y - x_{n+1}\rangle \le \mu_n\langle\tilde{w}_n, y - x_{n+1}\rangle \tag{3.4}$$

for all $y \in \mathcal{H}_n$. Since $\tilde{w}_n \in \partial_2 f(y_n, x_{n+1})$, we have

$$\langle\tilde{w}_n, y - x_{n+1}\rangle \le f(y_n, y) - f(y_n, x_{n+1}) \tag{3.5}$$

for all $y \in \mathcal{H}$. From (3.4) and (3.5), we get

$$\langle\rho_n - x_{n+1}, y - x_{n+1}\rangle \le \mu_n f(y_n, y) - \mu_n f(y_n, x_{n+1}) \tag{3.6}$$

for all $y \in \mathcal{H}_n$. Substituting $y = \varpi$ in (3.6), we obtain

$$\langle\rho_n - x_{n+1}, \varpi - x_{n+1}\rangle \le \mu_n f(y_n, \varpi) - \mu_n f(y_n, x_{n+1}). \tag{3.7}$$

Given $\varpi \in EP(f,C)$ imply that $f(\varpi, y_n) \ge 0$ and owing to the item $(\mathcal{C}1)$ gives that $f(y_n, \varpi) \le 0$. Thus, we obtain

$$\langle\rho_n - x_{n+1}, x_{n+1} - \varpi\rangle \ge \mu_n f(y_n, x_{n+1}). \tag{3.8}$$

Following the condition $(\mathcal{C}2)$, we have

$$f(y_n,x_{n+1}) \geq f(\rho_n,x_{n+1}) - f(\rho_n,y_n) - \ell_1 \| \rho_n - y_n \|^2 - \ell_2 \| y_n - x_{n+1} \|^2 . \tag{3.9}$$

Combining (3.8) and (3.9), we get

$$\langle \rho_n - x_{n+1}, x_{n+1} - \varpi \rangle \geq \mu_n f(\rho_n,x_{n+1}) - \mu_n f(\rho_n,y_n) - \ell_1\mu_n \| \rho_n - y_n \|^2 - \ell_2\mu_n \| y_n - x_{n+1} \|^2 . \tag{3.10}$$

By using the half-space definition, we have $\langle \rho_n - \mu_n w_n - y_n, x_{n+1} - y_n \rangle \leq 0$, which implies that

$$\langle \rho_n - y_n, x_{n+1} - y_n \rangle \leq \mu_n \langle w_n, x_{n+1} - y_n \rangle. \tag{3.11}$$

Since $w_n \in \partial_2 f(\rho_n, y_n)$, we obtain

$$\langle w_n, y - y_n \rangle \leq f(\rho_n, y) - f(\rho_n, y_n)$$

for all $y \in \mathcal{H}$. By replacing $y = x_{n+1}$, we obtain

$$\langle w_n, x_{n+1} - y_n \rangle \leq f(\rho_n, x_{n+1}) - f(\rho_n, y_n). \tag{3.12}$$

It follows from inequalities (3.11) and (3.12) that

$$\langle \rho_n - y_n, x_{n+1} - y_n \rangle \leq \mu_n f(\rho_n, x_{n+1}) - \mu_n f(\rho_n, y_n). \tag{3.13}$$

From (3.10) and (3.13), we have

$$\langle \rho_n - x_{n+1}, x_{n+1} - \varpi \rangle \geq \langle \rho_n - y_n, x_{n+1} - y_n \rangle - \ell_1\mu_n \| \rho_n - y_n \|^2 - \ell_2\mu_n \| y_n - x_{n+1} \|^2 . \tag{3.14}$$

Now, we obtain the following equalities:

$$\| \rho_n - \varpi \|^2 - \| x_{n+1} - \rho_n \|^2 - \| x_{n+1} - \varpi \|^2 = 2\langle \rho_n - x_{n+1}, x_{n+1} - \varpi \rangle$$

and

$$\| \rho_n - y_n \|^2 + \| x_{n+1} - y_n \|^2 - \| \rho_n - x_{n+1} \|^2 = 2\langle \rho_n - y_n, x_{n+1} - y_n \rangle.$$

Combining the above equalities with expression (3.14) finalizes the proof.

Lemma 3.3: *Assume that the items $(\mathcal{C}1)-(\mathcal{C}4)$ hold. If there is a subsequence $\{\rho_{n_k}\}$ of $\{\rho_n\}$ such that $\rho_{n_k} \rightharpoonup x^* \in \mathcal{H}$ and*

$$\lim_{k\to\infty} \| \rho_{n_k} - y_{n_k} \| = \lim_{k\to\infty} \| \rho_{n_k} - x_{n_k+1} \| = \lim_{k\to\infty} \| x_{n_k+1} - y_{n_k} \| = 0. \tag{3.15}$$

Then $x^* \in EP(f, C)$.

Proof. From $y_n \in C$, $\rho_{n_k} \rightharpoonup x^*$ and $\lim_{k\to\infty} \| \rho_{n_k} - y_{n_k} \| = 0$, we get $y_{n_k} \rightharpoonup x^* \in C$. This follows from $\lim_{k\to\infty} \| x_{n_k+1} - y_{n_k} \| = 0$ that the subsequence $\{x_{n_k+1}\}$ is bounded. For any $y \in \mathcal{H}_n$, using (3.6), (3.9) and (3.13), we have

$$\begin{aligned}
\mu_{n_k} f(y_{n_k}, y) &\geq \mu_{n_k} f(y_{n_k}, x_{n_k+1}) + \langle \rho_{n_k} - x_{n_k+1}, y - x_{n_k+1} \rangle \\
&\geq \mu_{n_k} f(\rho_{n_k}, x_{n_k+1}) - \mu_{n_k} f(\rho_{n_k}, y_{n_k}) - \ell_1\mu_{n_k} \| \rho_{n_k} - y_{n_k} \|^2 - \ell_2\mu_{n_k} \| y_{n_k} - x_{n_k+1} \|^2 \\
&\quad + \langle \rho_{n_k} - x_{n_k+1}, y - x_{n_k+1} \rangle \\
&\geq \langle \rho_{n_k} - y_{n_k}, x_{n_k+1} - y_{n_k} \rangle + \langle \rho_{n_k} - x_{n_k+1}, y - x_{n_k+1} \rangle - \ell_1\mu_{n_k} \| \rho_{n_k} - y_{n_k} \|^2 \\
&\quad - \ell_2\mu_{n_k} \| y_{n_k} - x_{n_k+1} \|^2 .
\end{aligned}$$

This implies by (3.15) and the boundedness of $\{x_{n_k+1}\}$ that the right hand side tends to zero. Due to $0<\mu_{n_k}\le\mu<\min\left\{\frac{1}{2\ell_1},\frac{1}{2\ell_2}\right\}$, the condition $(\mathcal{C}4)$, and $y_{n_k}\rightharpoonup x^*$, we obtain $0\le\limsup_{k\to\infty}f(y_{n_k},y)\le f(x^*,y)$ for all $y\in\mathcal{H}_n$. Since $C\subset\mathcal{H}_n$, we get $f(x^*,y)\ge 0$ for all $y\in C$, that is, $x^*\in EP(f,C)$.

With the above results we are now ready for the main convergence theorem.

Theorem 3.4: *Let the sequence $\{x_n\}$ generated due to Algorithm 1 and the items $(\mathcal{C}1)-(\mathcal{C}4)$ are satisfied. Then, $\{x_n\}$ converges strongly to $\xi=P_{EP(f,C)}\circ\nabla\phi(\xi)$.*

Proof. Let $\varpi\in EP(f,C)$. From $\lim_{n\to\infty}\frac{\delta_n}{\eta_n}\|x_n-x_{n-1}\|=0$, we get

$$\delta_n\|x_n-x_{n-1}\|\le\eta_n M_0 \tag{3.16}$$

for some $M_0>0$. Since $\nabla\phi$ is contraction with constant $\rho\in[0,1)$ and using (3.16), the following relation is obtained:

$$\begin{aligned}\|\rho_n-\varpi\|&\le\eta_n\|\nabla\phi(x_n)-\varpi\|+(1-\eta_n)\|x_n-\varpi\|+\delta_n\|x_n-x_{n-1}\|\\&\le\eta_n\|\nabla\phi(x_n)-\nabla\phi(\varpi)\|+\eta_n\|\nabla\phi(\varpi)-\varpi\|+(1-\eta_n)\|x_n-\varpi\|+\eta_n M_0\\&\le(1-\gamma_n)\|x_n-\varpi\|+\eta_n(\|\nabla\phi(\varpi)-\varpi\|+M_0)\\&=(1-\gamma_n)\|x_n-\varpi\|+\gamma_n M_1\\&\le\max\{\|x_n-\varpi\|,M_1\},\end{aligned}\tag{3.17}$$

where $\gamma_n=\eta_n(1-\rho)$ and $M_1=\frac{\|\nabla\phi(\varpi)-\varpi\|+M_0}{1-\rho}$. By $0<\mu_n\le\mu<\min\left\{\frac{1}{2\ell_1},\frac{1}{2\ell_2}\right\}$ with expression (3.3) implies that

$$\|x_{n+1}-\varpi\|\le\|\rho_n-\varpi\|. \tag{3.18}$$

This leads to a conclusion that $\|x_{n+1}-\varpi\|\le\max\{\|x_1-\varpi\|,M_1\}$ for any $n\in\mathbb{N}$. Consequently, the sequence $\{x_n\}$ is bounded. In addition, $\{\nabla\phi(x_n)\}$ is also bounded. Since f satisfies the conditions $(\mathcal{C}1)-(\mathcal{C}4)$, we have that the solution set $EP(f,C)$ is closed and convex, see [13]. Hence, $P_{EP(f,C)}\circ\nabla\phi$ is a ρ-contractive mapping. Now, we can uniquely find $\xi\in EP(f,C)$ with $\xi=P_{EP(f,C)}\circ\nabla\phi(\xi)$ due to the Banach fixed point theorem. By (2.1), we also get that for any $y\in EP(f,C)$,

$$\langle\nabla\phi(\xi)-\xi,y-\xi\rangle\le 0. \tag{3.19}$$

Now for each $n\in\mathbb{N}$, set $\Xi_n:=\|x_n-\xi\|^2$. Applying (3.17), we have

$$\begin{aligned}\|\rho_n-\xi\|^2&\le((1-\gamma_n)\|x_n-\xi\|+\gamma_n M_1)^2\\&=(1-\gamma_n)^2\Xi_n+\gamma_n(2M_1(1-\gamma_n)\|x_n-\xi\|+\gamma_n M_1^2)\\&\le\Xi_n+\gamma_n M_2\end{aligned}$$

for some $M_2>0$. This follows from (3.3) that

$$\begin{aligned}(1-2\ell_1\mu_n)\|\rho_n-y_n\|^2+(1-2\ell_2\mu_n)\|y_n-x_{n+1}\|^2&\le\|\rho_n-\xi\|^2-\Xi_{n+1}\\&\le\Xi_n-\Xi_{n+1}+\gamma_n M_2.\end{aligned}\tag{3.20}$$

Case a. We can find $N \in \mathbb{N}$ satisfying that for all $n \geq N$, the inequality $\Xi_{n+1} \leq \Xi_n$ holds. This together with the boundedness of $\{\Xi_n\}$, it is convergent. Due to the fact that $\lim_{n\to\infty}\gamma_n = 0$ and $0 < \mu_n \leq \mu < \min\left\{\frac{1}{2\ell_1}, \frac{1}{2\ell_2}\right\}$, and by (3.20),

$$\lim_{n\to\infty} \| \rho_n - y_n \| = \lim_{n\to\infty} \| y_n - x_{n+1} \| = 0, \tag{3.21}$$

which implies

$$\lim_{n\to\infty} \| \rho_n - x_{n+1} \| = 0. \tag{3.22}$$

From the definition of ρ_n, the inequality (3.16) and $\lim_{n\to\infty}\eta_n = 0$, we have

$$\begin{aligned} \| \rho_n - x_n \| &\leq \eta_n \| \nabla\phi(x_n) - x_n \| + \delta_n \| x_n - x_{n-1} \| \\ &\leq \eta_n (\| \nabla\phi(x_n) - x_n \| + M_0) \to 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{3.23}$$

This together with (3.22) implies that

$$\lim_{n\to\infty} \| x_{n+1} - x_n \| = 0. \tag{3.24}$$

Next observe that, for the reason that $\{x_n\}$ is bounded, there is $x^* \in \mathcal{H}$ such that $x_{n_k} \rightharpoonup x^*$ as $k \to \infty$ for some subsequence $\{x_{n_k}\}$ of $\{x_n\}$. By (3.23), we get $\rho_{n_k} \rightharpoonup x^*$ as $k \to \infty$. Then Lemma 3.3 together with (3.21) and (3.22) implies that $x^* \in EP(f, C)$. From (3.19), it is straightforward to show that

$$\limsup_{n\to\infty} \langle \nabla\phi(\xi) - \xi, x_n - \xi \rangle = \lim_{k\to\infty} \langle \nabla\phi(\xi) - \xi, x_{n_k} - \xi \rangle = \langle \nabla\phi(\xi) - \xi, x^* - \xi \rangle \leq 0.$$

This combining with (3.24) finds that

$$\limsup_{n\to\infty} \langle \nabla\phi(\xi) - \xi, x_{n+1} - \xi \rangle \leq \limsup_{n\to\infty} \langle \nabla\phi(\xi) - \xi, x_{n+1} - x_n \rangle + \limsup_{n\to\infty} \langle \nabla\phi(\xi) - \xi, x_n - \xi \rangle \leq 0. \tag{3.25}$$

Hence, from the assumption on $\nabla\phi$, (2.2) and (2.3), we obtain

$$\begin{aligned} \check{z}_{n+1} &\leq \| \rho_n - \xi \|^2 \\ &= \| \eta_n(\nabla\phi(x_n) - \nabla\phi(\xi)) + (1-\eta_n)(x_n - \xi) + \delta_n(x_n - x_{n-1}) + \eta_n(\nabla\phi(\xi) - \xi) \|^2 \\ &\leq \| \eta_n(\nabla\phi(x_n) - \nabla\phi(\xi)) + (1-\eta_n)(x_n - \xi) \|^2 + 2\langle \delta_n(x_n - x_{n-1}) + \eta_n(\nabla\phi(\xi) - \xi), x_{n+1} - \xi \rangle \\ &\leq \eta_n \| \nabla\phi(x_n) - \nabla\phi(\xi) \|^2 + (1-\eta_n)\Xi_n + 2\delta_n \langle x_n - x_{n-1}, x_{n+1} - \xi \rangle + 2\eta_n \langle \nabla\phi(\xi) - \xi, x_{n+1} - \xi \rangle \\ &\leq \eta_n \rho^2 \Xi_n + (1-\eta_n)\Xi_n + 2\delta_n \| x_n - x_{n-1} \| \| x_{n+1} - \xi \| + 2\eta_n \langle \nabla\phi(\xi) - \xi, x_{n+1} - \xi \rangle \\ &\leq \eta_n \rho \Xi_n + (1-\eta_n)\Xi_n + 2\eta_n \cdot \frac{\delta_n}{\eta_n} \| x_n - x_{n-1} \| \| x_{n+1} - \xi \| + 2\eta_n \langle \nabla\phi(\xi) - \xi, x_{n+1} - \xi \rangle \\ &\leq (1-\gamma_n)\Xi_n + \gamma_n \left[M_3 \frac{\delta_n}{\eta_n} \| x_n - x_{n-1} \| + \frac{2}{1-\rho} \langle \nabla\phi(\xi) - \xi, x_{n+1} - \xi \rangle \right] \end{aligned} \tag{3.26}$$

for some $M_3 > 0$. Applying this to the inequality (3.25) with Lemma 2.2, we can conclude that $\lim_{n\to\infty}\Xi_n = 0$.

Case b. We can find a subsequence $\{\Xi_{n_j}\}$ of $\{\Xi_n\}$ such that $\Xi_{n_j} < \Xi_{n_j+1}$ for all $j \in \mathbb{N}$. According to Lemma 2.3, the inequality $\check{z}_{\psi(n)} \leq \check{z}_{\psi(n)+1}$ is obtained, where $\psi : \mathbb{N} \to \mathbb{N}$ is defined by (2.4), and $n \geq n^*$ for some $n^* \in \mathbb{N}$. This implies, by (3.20), for all $n \geq n^*$, that

$$(1 - 2\ell_1\mu_{\psi(n)}) \| \rho_{\psi(n)} - y_{\psi(n)} \|^2 + (1 - 2\ell_2\mu_{\psi(n)}) \| y_{\psi(n)} - x_{\psi(n)+1} \|^2 \leq \Xi_{\psi(n)} - \Xi_{\psi(n)+1} + \gamma_{\psi(n)} M_2.$$

Similar as in Case a, since $\gamma_n \to 0$ as $n \to \infty$, we obtain

$$\lim_{n\to\infty} \| \rho_{\psi(n)} - y_{\psi(n)} \| = \lim_{n\to\infty} \| y_{\psi(n)} - x_{\psi(n)+1} \| = 0.$$

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

$$\limsup_{n\to\infty} \langle \nabla\phi(\xi) - \xi, x_{\psi(n)+1} - \xi \rangle \le 0.$$

Finally, from the inequality $\Xi_{\psi(n)} \le \Xi_{\psi(n)+1}$ and by (3.26), for all $n \ge n^*$, we obtain

$$\Xi_{\psi(n)+1} \le (1-\gamma_{\psi(n)})\Xi_{\psi(n)+1} + \gamma_{\psi(n)} \left[M_3 \frac{\delta_{\psi(n)}}{\eta_{\psi(n)}} \| x_{\psi(n)} - x_{\psi(n)-1} \| + \frac{2}{1-\rho} \langle \nabla\phi(\xi) - \xi, x_{\psi(n)+1} - \xi \rangle \right].$$

Some simple calculations yield

$$\Xi_{\psi(n)+1} \le M_3 \frac{\delta_{\psi(n)}}{\eta_{\psi(n)}} \| x_{\psi(n)} - x_{\psi(n)-1} \| + \frac{2}{1-\rho} \langle \nabla\phi(\xi) - \xi, x_{\psi(n)+1} - \xi \rangle.$$

This follows that $\limsup_{n\to\infty} \Xi_{\psi(n)+1} \le 0$. Thus, $\lim_{n\to\infty} \Xi_{\psi(n)+1} = 0$. In addition, by Lemma 2.3,

$$\lim_{n\to\infty} \Xi_n \le \lim_{n\to\infty} \Xi_{\psi(n)+1} = 0.$$

Therefore, we can conclude that $x_n \to \xi$ as $n \to \infty$.

4. Application to Educational Data Classification Problem

0.7 cm The educational dataset classification displayed in this application is pre-service teachers' digital proficiency level identifying as A1, A2, B1, B2 and C1. According to DigCompEdu framework [20], digital proficiency levels of prospective mathematics teachers were classified in this study: Newcomer (A1), Explorer (A2), Integrator (B1), Expert (B2), and Leader (C1).

We next give the concept of an extreme learning machine (ELM) [24] for applying our proposed algorithm to solve ..... prediction. We start by letting $\mathcal{U} := \{(x_n, b_n) : x_n \in \mathbb{R}^M, b_n \in \mathbb{R}^N, n = 1,2,\ldots,G\}$ as a training set of G distinct samples where x_n is an input training data and b_n is a target. For each output layer of ELM for single-hidden layer feed forward neural networks (SLFNs) with m hidden nodes and activation function A is

$$O_n = \sum_{j=1}^{m} (w_j A(a_j, c_j, x_n),$$

where a_j and c_j are parameters of weight and finally the bias, respectively. To find the optimal output weight w_j at the j-th hidden node, then the hidden layer output matrix A is generated as follows:

$$A = \begin{bmatrix} A(a_1, c_1, x_1) & \ldots & A(a_m, c_m, x_1) \\ \vdots & \ddots & \vdots \\ A(a_1, c_1, x_G) & \ldots & A(a_m, c_m, x_G) \end{bmatrix}.$$

For finding the optimal weight vector $w = [w_1^T, \ldots, w_m^T]^T$ which satisfies $Aw = B$ where $B = [t_1^T, \ldots, t_G^T]^T$ is the training target data. This problem can be solved by the least square problem when the *Moore-Penrose generalized inverse* of the matric A can not be fined easily. For ovoid overfitting in the machine learning, we use least square regularization. This problem can determine as the following convex minimization problem:

$$\min_{w\in\mathbb{R}^m} \{ \| Aw - B \|_2^2 + \lambda \| w \|_1 \}, \tag{4.1}$$

where λ is a regularization parameter. This problem is called the least absolute shrinkage and selection operator (LASSO) [25]. For applying our algorithms we set $f(u,v) = \langle A^T(Au - B), v - u\rangle$.

We use four evaluation metrics: Accuracy, Precision, Recall, and F1-score [26] as explained below for comparing the performance of the classification algorithms.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%. \tag{4.2}$$

$$Precision = \frac{TP}{TP + FP}. \tag{4.3}$$

$$Recall = \frac{TP}{TN + FN}. \tag{4.4}$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall}, \tag{4.5}$$

where these matrices gave True Negative (*TN*), False Positive (*FP*), False Negative (*FN*), and True Positive (*TP*). The multi-class cross entropy loss is used in multi-class classification by the form:

$$Loss = -\sum_{i=1}^{K} y^k \log \hat{y}^k, \tag{4.6}$$

where y^k is 0 or 1, indicating whether class label k is the correct classification and $\hat{y}^k$ is a probability of class y^k and K is the number of scalar values in the model output.

For starting our computation, we set the activation function as sigmoid, hidden nodes $m = 120$. We next start to find the suitable regularization parameter λ for our Algorithm 1 by setting $\mu_n = \frac{1}{\max(eigenvalue(A^T A))}$, $\rho = 0.5$, $\eta_n^* = \frac{1}{10n+1}$, $\delta_n^* = 0.5$ where

$$\delta_n = \begin{cases} \dfrac{1}{n \| x_n - x_{n-1} \|}, & \text{if } n \geq N, x_n \neq x_{n-1}, \\ \delta_n^*, & \text{otherwise,} \end{cases} \tag{4.7}$$

$$\eta_n = \begin{cases} \dfrac{1}{n}, & \text{if } n \geq N, \\ \eta_n^*, & \text{otherwise} \end{cases} \tag{4.8}$$

such that N is the number of iteration that we want to stop, and $\nabla\phi(x) = \rho x$, $\forall x \in \mathcal{H}$ with $\rho \in [0,1)$. The stopping criteria is the best accuracy of each case. The comparison of all cases with different parameters λ of Algorithm 1 are presented in Table 1.

From Table 1, we see that $\lambda = 10$ gets the highest accuracy, thus we choose $\lambda = 10$ for the next calculation. Next, we consider the different of the parameters μ_n when $\rho = 0.5$, $\eta_n^* = \frac{1}{10n+1}$ and $\delta_n^* = 0.5$. Then we obtain the following numerical results of different parameter μ_n.

From Table 2, we see that $\mu_n = \frac{0.999}{\max(eigenvalue(A^T A))}$ gets less number of iteration and the highest accuracy, thus we choose $\mu_n = \frac{0.999}{\max(eigenvalue(A^T A))}$ for the next calculation. Next, we consider

Table 1: Numerical results of different regularization parameters λ.

λ	Training Time	Iteration Number	Accuracy
.0001	0.0486	59	57.85
.01	0.0557	58	57.85
.1	0.0406	49	58.68
	0.0473	61	49.59
	0.0778	107	59.50

Table 2: Numerical results of different parameters μ_n.

μ_n	Training Time	Iteration Number	Accuracy
$\frac{0.1}{\max(eigenvalue(A^T A))}$	0.1127	154	57.85
$\frac{0.5}{\max(eigenvalue(A^T A))}$	0.0297	32	57.85
$\frac{0.99}{\max(eigenvalue(A^T A))}$	0.0431	34	57.87
$\frac{0.999}{\max(eigenvalue(A^T A))}$	0.0764	105	59.50
$\frac{1}{\max(eigenvalue(A^T A))}$	0.0778	107	59.50

Table 3: Numerical results of different parameters ρ.

ρ	Training Time	Iteration Number	Accuracy
0.1	0.1127	113	59.50
0.5	0.0297	105	59.50
0.9	0.0431	98	59.50
0.99	0.0764	96	59.50
0.999	0.0778	96	59.50

the different of the parameters ρ when $\lambda = 10$, $\eta_n^* = \frac{1}{10n+1}$ and $\delta_n^* = 0.5$. Then we obtain the following numerical results of different parameter ρ.

From Table 3, we see that $\rho = 0.99$ gets less number of iteration and the highest accuracy, thus we choose $\rho = 0.99$ for the next calculation. Next, we consider the different of the parameters η_n^* when $\lambda = 10$, $\mu_n = \frac{0.999}{\max(eigenvalue(A^T A))}$ and $\delta_n^* = 0.5$. Then we obtain the following numerical results of different parameter η_n^*.

From Table 4, we see that $\eta_n^* = \frac{1}{100n+1}$ gets less number of iteration and training time, thus we choose $\eta_n^* = \frac{1}{100n+1}$ for the next calculation. Next, we consider the different of the parameters

Table 4: Numerical results of different parameters η_n^*.

η_n^*	Training Time	Iteration Number	Accuracy
$\frac{1}{n+1}$	0.0741	98	59.50
$\frac{1}{10n+1}$	0.0745	96	59.50
$\frac{1}{50n+1}$	0.0745	96	59.50
$\frac{1}{100n+1}$	0.0729	96	59.50
$\frac{1}{1000n+1}$	0.0767	96	59.50

Table 5: Numerical results of different parameters δ_n^*.

δ_n^*	Training Time	Iteration Number	Accuracy
0.2	0.1103	152	59.50
0.5	0.0729	96	59.50
0.9	0.0254	26	57.85
$\frac{10}{\|x_n - x_{n-1}\| + 10}$	0.0508	65	62.81
$\frac{1}{\|x_n - x_{n-1}\| + n^3}$	0.1316	191	59.50

δ_n^* when $\lambda = 10$, $\mu_n = \frac{0.999}{\max(eigenvalue(A^T A))}$ and $\rho = 0.99$. Then we obtain the following numerical results of different parameter δ_n^*.

From Table 5, we see that $\delta_n^* = \frac{10}{\| x_n - x_{n-1} \| + 10}$ gets the highest accuracy. We next show the performance of our Algorithm 1 compare with the other existing algorithms (1.2), (1.3), (1.5) and (1.7).

Table 6 shows that our algorithm gets the highest precision, recall, F1-score, and accuracy efficiency. Even though it has a higher number of iterations than Algorithm (1.7) but it has less training number than that. It has the highest probability of correctly classifying pre-service mathematics teachers' technology integrated competency level compared to algorithms examinations. We present the training and validation loss with the accuracy of training to show that our algorithm has no overfitting in the training dataset.

From Figures 1–2, we see that our model from Algorithm 3 by the suitable parameters in Table 1–6 gets good fitting model that is the measure of a machine learning model generalizes well to similar data to that on.

As a result, we implemented a modified extragradient method with an inertial extrapolation step and viscosity-type method to solve equilibrium problems of the pseudomonotone bifunction operator

Table 6: The performance of our Algorithm 3 comparing with the other exiting algorithms.

	Iter. No.	Training Time	Precision	Recall	F1-score	Accuracy
Algorithm (1.2)	190	0.1445	0.6956	0.6538	0.5798	59.50
Algorithm (1.3)	192	0.1462	0.6956	0.6538	0.5798	59.50
Algorithm (1.5)	191	0.6159	0.6956	0.6538	0.5798	59.50
Algorithm (1.7)	58	0.1385	0.6133	0.6786	0.6322	61.16
Algorithm 1	65	0.0508	0.7053	0.6923	0.6292	62.81

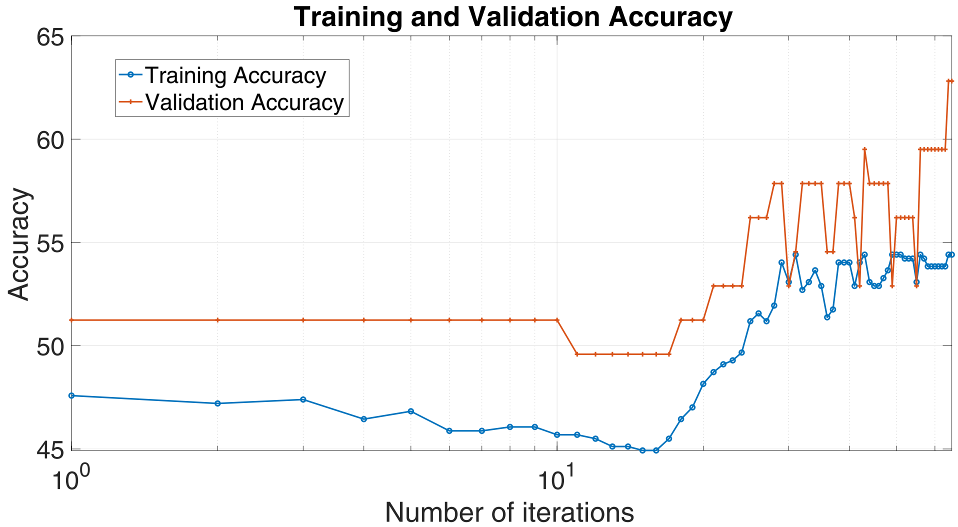

Figure 1: Accuracy plots of the iteration of Algorithm 1.

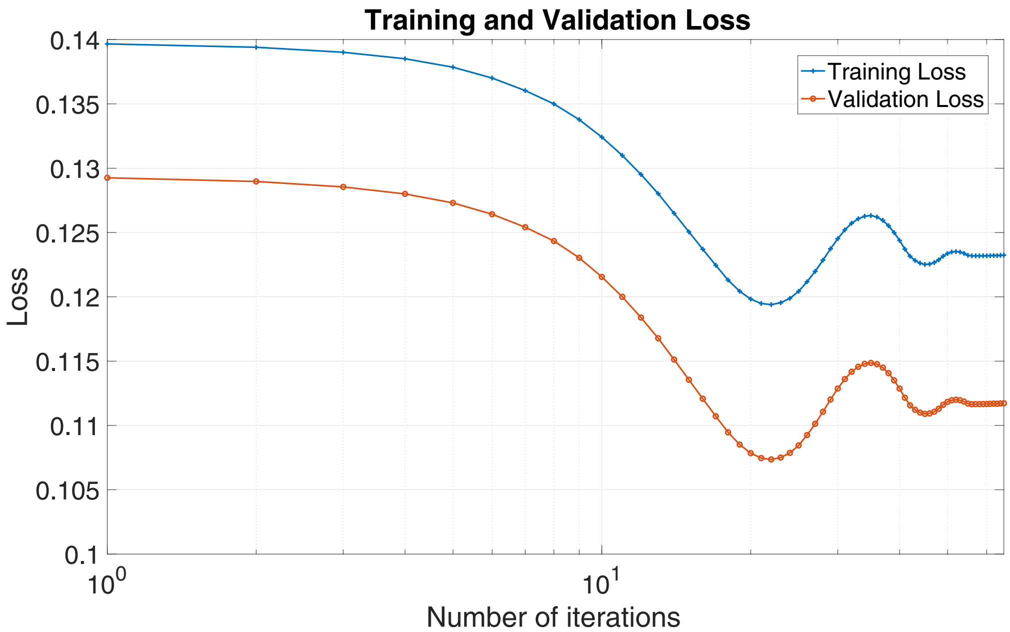

Figure 2: Loss plots of the iteration of Algorithm 1.

to an educational dataset of 474 instances containing nine attributes, including major; gender; type of supplementary digital technology training courses (short, medium, and long format); grade point of Information Technology for Learning course; grade point of Innovation and Information Technology course; grade point of Digital Technology for teaching mathematics, GPA, and digital proficiency level. The accuracy of classification achieved by the proposed machine learning algorithm. It was found that 62.81% of the dataset was been classified accurately with our algorithm although the number of iterations compared to Algorithm 1.7 may be higher.

5. Conclusion and Discussion

0.7cm This paper proposes a modified extragradient method as a machine learning-based model to predict the digital proficiency level in mathematics classrooms of pre-service teachers, using their major; gender; type of supplementary digital technology training courses; grade point of Information Technology for Learning course; grade point of Innovation and Information Technology course; grade point of Digital Technology for teaching mathematics course, and GPA as the primary data. To predict the pre-service teachers' digital proficiency level, the performance of several machine learning methods was computed and compared.

According to the findings, the proposed method attained a classification accuracy of 62.81%, higher than other algorithms. Subsequently, it can be said that grade points of existing courses related to the digital technology of pre-service teachers who deal with mathematics classrooms are critical predictors to be used to predict digital proficiency level. The result was compared with the studies that predicted digital and technological knowledge level of pre-service teachers. Leoste et al. [27] predicted digital skills of early childhood teachers in Estonia to be successful in teaching with digital technology. The duration type of a supplementary digital technology training course affected to the teachers' digital skills. The findings of their study imply that a short-term training course would be practical for leading teachers with beginner-level digital competencies to the expert level, while a long-term training course might result in a more significant proportion of teachers with top-level digital skills.

Moreover, the result presented that the Information Technology for Learning course grade points affected pre-service teachers' digital proficiency levels. While the Information Technology for Learning course aims to develop the technology knowledge of pre-service teachers, this result is consistent with Trainin et al. [28]. They found that technology knowledge significantly predicted pre-service teachers' instructional change with technology integration.

Angers and Machtmes [29] predicted the technology-integrating skills of middle school exemplary teachers that a perspective about computer technology as a tool for teaching and learning affected the integrating skills. This is relevant to the Innovation and Information Technology course grade point attribute. Since the Innovation and Information Technology course included content related to a perspective of using computer technology as a tool for designing practical lessons, they found that it has a significant effect on teachers' technological integrating skills.

Furthermore, according to our result, the grade point of Digital Technology for teaching mathematics course affected the digital proficiency level. This course aims to develop technology-enhanced communities and core content of teachers' pedagogical reasoning using technology in mathematics classrooms. Thus, our result is coherent with Guzey and Roehrig [30]. They concluded their study that the formation of Technological Pedagogical and Content Knowledge (TPACK), which involves digital proficiency level, was strongly tied to teachers' pedagogical reasoning, and Technology Enhanced Community (TEC) pushed teachers to assess their pedagogical reasoning and practices critically.

Consequently, the digital proficiency level of pre-service teachers in charge of mathematics classrooms was predicted using our proposed modified extragradient method. The findings demonstrate that our machine learning algorithms accurately predict pre-service teachers' digital proficiency levels. The results of this study may assist teacher educators in identifying pre-service teachers with below-average or above-average academic motivation. Later, for instance, teachers may pair students with below-average educational drive with students with above-average digital proficiency and urge them to engage in groups or projects. This may increase the pre-service teachers' motivation and ensure their active learning process engagement. In addition, such data-driven research could aid higher education in building a framework for learning analytics and contribute to the decision-making processes of pre-service teachers.

Future research may include more input attributes and machine learning methods in the modeling procedure. In addition, it is vital to leverage the efficacy of many approaches to analyze pre-service teachers' learning habits, solve their issues in teaching, and enhance the educational environment.

Acknowledgments

We would like to thank Lampang Rajabhat University for supporting the raw dataset to be analyzed in our main result of this paper. Watcharaporn Cholamjiak would like to thank National Research Council of Thailand and University of Phayao (N42A650334); Thailand Science Research and Innovation, University of Phayao (FF66-UoE). Nipa Jun-on would like to thank the revenue budget in 2023, Lampang Rajabhat University.

References

[1] Galindo-Dominguez, H., and Bezanilla, M.J., *Digital Competence in the Training of Pre-service Teachers: Perceptions of Students in the Degrees of Early Childhood Education and Primary Education*. J. Dig. Learn. Teach. Educ. 37, (2021), 262–278.

[2] UNICEF Thailand. Available online: https://www.unicef.org/thailand/stories/school-reopening-how-teachers-and-students-are-adjusting-new-normal-thailand (last accessed on 30 July 2022).

[3] Calder, N., Jafri, M., and Guo, L., *Mathematics Education Students' Experiences During Lockdown: Managing Collaboration in eLearning*. Educ. Sci. 11, (2021), 191.

[4] Perry, T., Indon, M., and Cordingley, P., *Remote and Blended Teacher Education: A Rapid Review*. Educ. Sci. 11, (2021), 453.

[5] Ahshan, R. *A Framework of Implementing Strategies for Active Student Engagement in Remote/Online Teaching and Learning during the COVID-19 Pandemic*. Educ. Sci. 11, (2021), 483.

[6] Bonafini, F.C., and Lee, Y., *Investigating Prospective Teachers' TPACK and Their Use of Mathematical Action Technologies as They Create Screencast Video Lessons on iPads*. TechTrends. 65, (2021), 303–319.

[7] Hill, H.C., Rowan, B., and Ball, D.L., *Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement*. Am. Educ. Res. J. 42, (2005), 371–406.

[8] Khairiree, K., *Online Learning and Augmented Reality: Enhancing Students Learn Transformation Geometry during the Covid-19 Pandemic. In proceedings of the 25th Asian Technology Conference in Mathematics*, Virtual, Online, (2020), 304–313.

[9] Barlovits, S., Jablonski, S., Lazaro, C., Ludwig, M., Recio, T., *Teaching from A Distance—Math Lessons during COVID-19 in Germany and Spain*. Educ. Sci. 11, (2021), 406.

[10] Chirinda, B., Ndlovu, M., Spangenberg, E., *Teaching Mathematics during the COVID-19 Lockdown in A Context of Historical Disadvantage*. Educ. Sci. 11, (2021), 177.

[11] Muu, L. D., Oettli, W., *Convergence of an Adaptive Penalty Scheme for Finding Constrained Equilibria*. Nonlinear Anal. Theory Methods Appl. 18, (1992), 1159–1166.

[12] Korpelevich, G., *The Extragradient Method for Finding Saddle Points and Other Problems*. Matecon. 12, (1976), 747–756.

[13] Tran, D. Q., Dung, M. L., Nguyen, V. H., *Extragradient Algorithms Extended to Equilibrium Problems*. Optimization 57, (2008), 749–776.

[14] Van Hieu, D., *Halpern Subgradient Extragradient Method Extended to Equilibrium Problems*. Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas 111, (2017), 823–840.

[15] Kraikaew, R., Saejung, S., *Strong Convergence of the Halpern Subgradient Extragradient Method for Solving Variational Inequalities in Hilbert Spaces*. J. Optim. Theory Appl. 163, (2014), 399–412.

[16] Moudafi, A., *Viscosity Approximation Methods for Fixed-points Problems*. J. Math. Anal. Appl., 241, (2000), 46–55.

[17] Muangchoo, K., *A New Strongly Convergent Algorithm to Solve Pseudomonotone Equilibrium Problems in a Real Hilbert Space*. J. Math. Computer Sci. 24, (2022), 308–322.

[18] Polyak, B.T., *Some Methods of Speeding Up the Convergence of Iteration Methods*. USSR Comput. Math. Math. Phys. 4, (1964), 1–17.

[19] Shehu, Y., Izuchukwu, C., Yao, J. C., Qin, X., *Strongly Convergent Inertial Extragradient Type Methods for Equilibrium Problems*. Appl. Anal. 2021, 1–29.

[20] European Commission. Available online at: https://educators-go-digital.jrc.ec.europa.eu/ (accessed March 9, 2022).

[21] Tiel, J. V., *Convex Analysis: An Introductory Text*, 1st ed., Wiley: New York, NY, (1984).

[22] Xu, H. K., *Iterative Algorithms for Nonlinear Operators*. J Lond Math Soc. 66, (2002), 240–256.

[23] Maingé, P. E., *Strong Convergence of Projected Subgradient Methods for Nonsmooth and Nonstrictly Convex Minimization*. Set-Valued Anal. 16, (2008), 899–912.

[24] Huang,G. B., Zhu, Q. Y., Siew, C. K., *Extreme Learning Machine: Theory and Applications*. Neurocomputing, (2006), 489–501.

[25] Tibshirani, R., *Regression Shrinkage and Selection via the Lasso*. J. R. Stat. Soc. Series B. Stat. Methodol. 58, (1996), 267–288.

[26] Han, J., Kamber, M., Pei, J., *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufman Publishers: Waltham, MA, (2012), 978.

[27] Leoste, J., Lavicza, Z., Fenyvesi, K., Tuul, M., Oun, T., *Enhancing Digital Skills of Early Childhood Teachers Through Online Science, Technology, Engineering, Art, Math Training Programs in Estonia*. Front. Educ. 7, (2022), 894142.

[28] Trainin, G., Friedrich, L., Deng, Q., *The Impact of a Teacher Education Program Redesign on Technology Integration in Elementary Preservice teachers*. CITE 18, (2018), 692–721.

[29] Angers, J., Machtmes, K., *An Ethnographic-Case Study of Beliefs, Context Factors, and Practices of Teachers Integrating*, Qual. Rep. 10, (2005), 771–794.

[30] Guzey, S. S., Roehrig, G. H., *Teaching Science with Technology: Case Studies of Science Teachers' Development of Technology, Pedagogy, and Content Knowledge*. CITE 9, (2009), 25–45.