Two inertial projective Mann forward-backward algorithm for variational inclusion problems and application to stroke prediction

PRONPAT PEEYADA, PONKAMON KITISAK, WATCHARAPORN CHOLAMJIAK and
DAMRONGSAK YAMBANGWAI

ABSTRACT. This paper presents a two inertial technique with a projection Mann forward-backward splitting algorithm for solving variational inclusion problems that exhibit weak convergence under suitable conditions in Hilbert spaces. Furthermore, we provide a numerical example in infinitely dimensional spaces to support the main result. Finally, we provide an application for data classification using an extreme learning machine. According to data provided by the World Health Organization (WHO), stroke is recognized as the predominant contributor to mortality and disability on a global scale. To appraise the efficacy of our algorithm, we procured a dependable dataset for stroke prediction from the Kaggle website. The best algorithm that performed this task is ours compared to other machine learning methods.

1. INTRODUCTION

In the past decade, extensive scholarly efforts have been dedicated to developing comprehensive problem formulations in optimization. It all started with [8], where an equilibrium problem was considered. Subsequently, the research landscape expanded significantly, evident from the extensive list of references [3, 7, 11, 17].

In recent years, significant research has been dedicated to the study of inclusion problems, which extend and generalize the concept of equilibrium problems. These inclusion problems cover various problem classes across various academic disciplines. They encompass topics such as variational inequalities, fixed point and coincidence point problems, complementarity problems, and the Nash equilibrium problem, among other notable areas [19, 27]. Noteworthy scholarly works, such as the references [2, 9, 10], specifically focus on a distinct subset of inclusion problems known as variational inclusion problems. These problems are primarily concerned with identifying and characterizing the zeros of maximal monotone mappings.

Throughout this paper, we consistently assume that H is a real Hilbert space with inner product $\langle ., . \rangle$ and norm $\|.\|$. Let $\mathsf{G} : H \to 2^H$ be a multi-valued monotone operator and $\mathsf{F} : H \to H$ be a single-valued operator. We consider the following variational inclusion problem: find a point $\hat{x} \in H$ such that

$$(1.1) \qquad\qquad 0 \in (\mathsf{F} + \mathsf{G})\hat{x}.$$

We denote the solution set of (1.1). The variational inclusion problem represents a substantial progression from the variational inequality problem within the field. By effectively reformulating various nonlinear problems such as saddle point problems, minimization problems, and split feasibility problems as variational inclusion problems, significant advancements are made. These transformations carry significant implications across a range of domains, including signal processing, neural networks, medical image reconstruction, machine learning, and data mining (see [1, 21, 25]).

An increasing number of scholars are actively researching diverse methodologies to tackle the intricate variational inclusion problem effectively. The forward-backward splitting method has gained prominence in this scholarly domain. The following exposition outlines this approach.

$$(1.2) \qquad x^{k+1} = J_r^{\mathsf{G}}(x^k - r\mathsf{F}x^k), k \geq 1,$$

where $J_r^{\mathsf{G}} = (I+r\mathsf{G})^{-1}$ with $r > 0$. Furthermore, researchers have modified these methods by increasing their versatility through relaxation techniques and improving their acceleration using inertial techniques.

Alvarez and Attouch [4] extended the heavy ball method to encompass a broader context involving a general maximal monotone operator. This extension was achieved by strategically integrating the proximal point algorithm framework, resulting in the creation of the advanced inertial proximal point algorithm (IPA), formally represented as:

$$(1.3) \qquad \begin{cases} y^k = x^k + \theta^k(x^k - x^{k-1}), \\ x^{k+1} = J_{r^k}^{\mathsf{G}}(y^k), k \geq 1. \end{cases}$$

They proved that under the specified condition.

$$(1.4) \qquad \sum_{k=1}^{\infty} \theta^k \|x^k - x^{k-1}\| < \infty,$$

the algorithm (1.3) weakly converges to a zero of G. An additional single-valued, cocoercive, and Lipschitz continuous operator, denoted as F, into the inertial proximal point algorithm, known as the splitting inertial proximal algorithm (SIPA) has been introduced by Moudafi and Oliny [23].

$$(1.5) \qquad \begin{cases} y^k = x^k + \theta^k(x^k - x^{k-1}), \\ x^{k+1} = J_{r^k}^{\mathsf{G}}(y^k - r^k\mathsf{F}x^k), k \geq 1. \end{cases}$$

They achieved a weak convergence result using the algorithm (1.5) under the same condition (1.4) as specified in [4]. As noted in [18], algorithm (1.5) does not use the format of a forward-backward splitting algorithm because operator F is still evaluated on the point x^k when $\theta^k > 0$.

Recently, Iyioa and Shehu [15] considered the second order dynamical system. Consequently, we can express the discretization of the system as follows:

$$(1.6) \qquad \begin{cases} y^k = x^k + \theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} + x^{k-2}), \\ x^{k+1} = (1 - \alpha^k)y^k + \alpha^k J_{r^k}^{\mathsf{G}}(y^k), \end{cases}$$

which is called the two-step inertial proximal point algorithm (TSIPA). Where $r^k > 0$, θ^k and δ^k satisfy some conditions.

Inspired and motivated by the aforementioned studies, we propose a two inertial technique with a projection Mann forward-backward splitting algorithm to solve variational inclusion problems in a real Hilbert spaces. Under some standard and mild conditions,

we obtain a weak convergence result for the proposed algorithm and obtain a consequent result. In addition, we provide a numerical example in infinitely dimensional spaces to support the main result. Finally, we also provide an application to predict the stroke dataset by using our proposed algorithm in an extreme learning machine. To ascertain which classification model predicts the dataset more accurately, we calculate and compare a range of performance metrics such as accuracy, precision, recall, and F1-score. The results revealed that the proposed algorithm has better efficiency in handling classification problems.

2. PRELIMINARIES

This section will rely on the following key lemmas to support our main findings.

Definition 2.1. Let $\mathsf{G} : H \to 2^H$ be a multi-valued operator. Then G is said to be
(i) monotone if $\langle p - q, x - y \rangle \geq 0$, for all $(x, p), (y, q) \in gra(\mathsf{G})$ (the graph of operator G),
(ii) maximal monotone if G dose not admit proper monotone extension.

Definition 2.2. A operator $\mathsf{F} : H \to H$ is said to be
(i) firmly nonexpansive if for all $x, y \in H$,

$$\langle x - y, \mathsf{F}x - \mathsf{F}y \rangle \geq \|\mathsf{F}x - \mathsf{F}y\|^2,$$

(ii) η-inverse strongly monotone if $\eta \mathsf{F}$ is firmly nonexpansive when $\eta > 0$,
(iii) L-Lipschitz continuous if there is $L > 0$ such that for all $x, y \in H$,

$$\|\mathsf{F}x - \mathsf{F}y\| \leq L\|x - y\|,$$

(iv) nonexpansive if for all $x, y \in H$, $\|\mathsf{F}x - \mathsf{F}y\| \leq \|x - y\|$.

According to the definition, it is clear that every η-cocoercive mapping is monotone and $\frac{1}{\eta}$-Lipschitz continuous. It is widely acknowledged that when $\mathsf{G} : H \to 2^H$ is a multi-valued maximal monotone operator and $r > 0$, the operator $J_r^{\mathsf{G}} := (I + r\mathsf{G})^{-1}$ represents a single-valued firmly nonexpansive mapping [22].

Lemma 2.1. [12] *Let* $\mathsf{F} : H \to H$ *be a nonexpansive mapping such that* $Fix(\mathsf{F}) \neq \emptyset$. *If there exists a sequence* $\{x^k\}$ *in* H *such that* $x^k \rightharpoonup x \in H$ *and* $\|x^k - \mathsf{F}x^k\| \to 0$, *then* $x \in Fix(\mathsf{F})$.

Lemma 2.2. [16] *Let* $\mathsf{F} : H \to H$ *be* η-*inverse strongly monotone and* $\mathsf{G} : H \to 2^H$ *a maximal monotone operator. Then, we have*
(i) for $r > 0$, $Fix(J_r^{\mathsf{G}}(I - r\mathsf{F})) = (\mathsf{F} + \mathsf{G})^{-1}(0)$,
(ii) for $0 < r < \bar{r}$ *and* $x \in H$, $\|x - J_r^{\mathsf{G}}(I - r\mathsf{F})x\| \leq 2\|x - J_{\bar{r}}^{\mathsf{G}}(I - \bar{r}\mathsf{F})x\|$

Lemma 2.3. [6] *Let* Γ *be a nonempty set of* H *and* $\{x^k\}$ *be a sequence in* H. *Assume that the following conditions hold.*
(i) for every $x \in \Gamma, \{\|x^k - x\|\}$ *converges.*
(ii) Every weak sequence cluster point of $\{x^k\}$ *belongs to* Γ.
Then $\{x^k\}$ *weakly converges to an element in* Γ.

Lemma 2.4. [5] *Let* $\{a^k\}$ *and* $\{b^k\}$ *be a nonnegative sequences of real number satisfying* $\sum_{k=1}^{\infty} b^k < \infty$ *and* $a^{k+1} \leq a^k + b^k$. *Then,* $\{a^k\}$ *is a convergent sequences.*

3. Main results

In the section, we suppose that that E is a nonempty closed and convex subset of H. Let $\mathsf{F} : H \to H$ be η-inverse strongly monotone and $\mathsf{G} : H \to 2^H$ a maximal monotone operator such that $(\mathsf{F} + \mathsf{G})^{-1}(0) \cap E \neq \emptyset$. We present our proposed algorithm for solving variational inclusion problem (1.1) and prove its weak convergence as the following Algorithm 3.0.1.

Algorithm 3.0.1. *Initialization: Let* $x^{-1}, x^0, x^1 \in H, \{\alpha^k\} \subset (a,b) \subset (0,1], \{r^k\} \subset (c,d) \subset (0,2\eta), \{\theta^k\}, \{\delta^k\} \subset (-\infty, \infty)$ *satisfy the following conditions:*

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

Step 1. Compute

$$y^k = x^k + \theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2}),$$

Step 2. Compute

$$x^{k+1} = P_E((1 - \alpha^k)x^k + \alpha^k J^k y^k),$$

where $J^k = J_{r^k}^{\mathsf{G}}\left(I - r^k \mathsf{F}\right)$. *Set* $k = k + 1$ *and return to **Step 1**.*

Theorem 3.1. *The sequence* $\{x^k\}$ *generated by Algorithm 3.0.1 converges weakly to an element in* $(\mathsf{F} + \mathsf{G})^{-1}(0) \cap E$.

Proof. Let $x^* \in (\mathsf{F} + \mathsf{G})^{-1}(0) \cap E$. Since $\{r^k\} \subset (0,2\eta)$, J^k is nonexpansive mapping, and F is η-inverse strongly monotone, we have

$$
\begin{aligned}
\|x^{k+1} - x^*\| &= \|P_E((1 - \alpha^k)x^k + \alpha^k J^k y^k) - x^*\| \\
&\leq (1 - \alpha^k)\|x^k - x^*\| + \alpha^k\|J^k y^k - x^*\| \\
&\leq (1 - \alpha^k)\|x^k - x^*\| + \alpha^k\|x^k + \theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2}) - x^*\| \\
&\leq (1 - \alpha^k)\|x^k - x^*\| + \alpha^k\|x^k - x^*\| + \alpha^k|\theta^k|\|x^k - x^{k-1}\| \\
&\quad + \alpha^k|\delta^k|\|x^{k-1} - x^{k-2}\| \\
(3.7) \quad &= \|x^k - x^*\| + \alpha^k|\theta^k|\|x^k - x^{k-1}\| + \alpha^k|\delta^k|\|x^{k-1} - x^{k-2}\|.
\end{aligned}
$$

By the conditions of $\{\theta^k\}$, $\{\delta^k\}$, it follows from Lemma 2.4, we obtain $\lim\limits_{k \to \infty} \|x^k - x^*\|$ exists. Since $J^{\mathsf{G}}_{r^k}$ is a firmly nonexpansive mapping then we have

$$
\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|P_E((1-\alpha^k)x^k + \alpha^k J^k y^k) - x^*\|^2 \\
&\leq (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k\|J^k y^k - x^*\|^2 \\
&\leq (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k[\|y^k - r^k\mathsf{F}y^k - x^* + r^k\mathsf{F}x^*\|^2 \\
&\quad -\|y^k - r^k\mathsf{F}y^k - J^k y^k - x^* + r^k\mathsf{F}x^* + J^k x^*\|^2] \\
&= (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k[\|y^k - x^* - r^k(\mathsf{F}y^k + \mathsf{F}x^*)\|^2 \\
&\quad -\|y^k - x^* - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2] \\
&= (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k[\|y^k - x^*\|^2 - 2r^k\langle y^k - x^*, \mathsf{F}y^k + \mathsf{F}x^*\rangle \\
&\quad +(r^k)^2\|\mathsf{F}y^k - \mathsf{F}x^*\|^2] - \alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2 \\
&\leq (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k\|y^k - x^*\|^2 - \alpha^k(r^k)^2\|Fy^k - Fx^*\|^2 \\
&\quad -2\eta\alpha^k r^k\|\mathsf{F}y^k - \mathsf{F}x^*\|^2 - \alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2 \\
&= (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k\|x^k + \theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2})^2 \\
&\quad -\alpha^k r^k(2\eta - r^k)\|\mathsf{F}y^k - \mathsf{F}x^*\|^2 - \alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2 \\
&\leq (1-\alpha^k)\|x^k - x^*\|^2 + \alpha^k\|x^k - x^*\|^2 + 2\alpha^k\langle\theta^k(x^k - x^{k-1}) \\
&\quad +\delta^k(x^{k-1} - x^{k-2}), y^k - x^*\rangle - \alpha^k r^k(2\eta - r^k)\|\mathsf{F}y^k - \mathsf{F}x^*\|^2 \\
&\quad -\alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2 \\
&= \|x^k - x^*\|^2 + 2\alpha^k\langle\theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2}), y^k - x^*\rangle \\
&\quad -\alpha^k r^k(2\eta - r^k)\|\mathsf{F}y^k - \mathsf{F}x^*\|^2 - \alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 + 2\alpha^k\langle\theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2}), y^k - x^*\rangle \\
&\quad \geq \alpha^k r^k(2\eta - r^k)\|\mathsf{F}y^k - \mathsf{F}x^*\|^2 + \alpha^k\|y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*) - J^k y^k\|^2.
\end{aligned}
\tag{3.8}
$$

Again by the conditions of the sequences $\{\theta^k\}$, $\{\delta^k\}$, conditions $(i) - (ii)$, and (3.8), we have

$$
\liminf_{k \to \infty} \|\mathsf{F}y^k - \mathsf{F}x^*\| = \liminf_{k \to \infty} \|y^k - J^k y^k - r^k(\mathsf{F}y^k - \mathsf{F}x^*)\| = 0.
$$

This gives, by the triangle inequality, that

$$
\liminf_{k \to \infty} \|y^k - J^k y^k\| = 0. \tag{3.9}
$$

Since $\liminf\limits_{k \to \infty} r^k > 0$, there is $r > 0$ such that $r^k > r$. Lemma 2.2 (ii), we obtain

$$
\lim_{k \to \infty} \|y^k - J^{\mathsf{G}}_r(I - r\mathsf{F})y^k\| \leq \lim_{k \to \infty} \|y^k - J^k y^k\| = 0.
$$

On the other hand,

$$
\|y^k - x^k\| \leq |\theta^k|\|x^k - x^{k-1}\| + |\delta^k|\|x^{k-1} - x^{k-2}\|. \tag{3.10}
$$

Since $\{x^k\}$ is bounded, we can let x^* be a week sequential cluster point of $\{x^k\}$. By (3.10), we have x^* is also a weak sequential cluster point of $\{y^k\}$. Using Lemma 2.2 (i) and Lemma 2.1, we get that $x^* \in Fix(J^{\mathsf{G}}_r(I - r\mathsf{F})) = (\mathsf{F} + \mathsf{G})^{-1}(0) \cap E$. Since x^k is a sequence in E and E is closed, it follows that $x^* \in (\mathsf{F} + \mathsf{G})^{-1}(0) \cap E$. By utilizing Opial's Lemma (Lemma 2.3), we can obtain that $\{x^k\}$ weakly converges to an element $(\mathsf{F}+\mathsf{G})^{-1}(0) \cap E$. $\quad\square$

Remark 3.1. Our Algorithm 3.0.1 can be reduced to following algorithms;

(i) if $\delta^k = 0$, we have

$$
\begin{aligned}
y^k &= x^k + \theta^k(x^k - x^{k-1}), \\
x^{k+1} &= P_E((1 - \alpha^k)x^k + \alpha^k J^k y^k),
\end{aligned}
$$

(ii) if $\theta^k = 0$,

$$
\begin{aligned}
y^k &= x^k + \delta^k(x^{k-1} - x^{k-2}), \\
x^{k+1} &= P_E((1 - \alpha^k)x^k + \alpha^k J^k y^k),
\end{aligned}
$$

(iii) if $\delta^k = 0$ and $\theta^k = 0$,

$$
\begin{aligned}
y^k &= x^k, \\
x^{k+1} &= P_E((1 - \alpha^k)x^k + \alpha^k J^k y^k),
\end{aligned}
$$

(iv) if $\alpha^k = 1$,

$$
\begin{aligned}
y^k &= x^k + \theta^k(x^k - x^{k-1}) + \delta^k(x^{k-1} - x^{k-2}), \\
x^{k+1} &= P_E(J^k y^k),
\end{aligned}
$$

(v) if $\delta^k = 0$ and $\alpha^k = 1$,

$$
\begin{aligned}
y^k &= x^k + \theta^k(x^k - x^{k-1}), \\
x^{k+1} &= P_E(J^k y^k),
\end{aligned}
$$

(vi) if $\theta^k = 0$ and $\alpha^k = 1$,

$$
\begin{aligned}
y^k &= x^k + \delta^k(x^{k-1} - x^{k-2}), \\
x^{k+1} &= P_E(J^k y^k).
\end{aligned}
$$

For supporting our main theorem, we now give an example in infinitely dimensional spaces $L_2[0,1]$ such that $\|.\|$ is L_2-norm defined by $\|x\| = \sqrt{\int_0^1 |x(t)|^2 dt}$ where $x(t) \in L_2[0,1]$.

Example 3.1. Let $H = L_2[0,1]$ and $E = \{x(t) \in L_2[0,1] : \int_0^1 x(t)dt < \infty\}$. Define mappings as follow:

(i) multi-valued monotone operator $\mathsf{G} : L_2[0,1] \to L_2[0,1]$ by $\mathsf{G}x(t) = 2x(t)$, $\forall x(t) \in L_2[0,1]$;

(ii) single-valued operator $\mathsf{F} : L_2[0,1] \to L_2[0,1]$ by $\mathsf{F}x(t) = 3x(t)$, $\forall x(t) \in L_2[0,1]$.

We can choose $x^{-1} = \frac{t+\sin(t)}{2}$, $x^0 = \frac{\sin(t)}{2}$ and $x^1 = \sin(t)$. We use the Cauchy error $\|x^k - x^{k-1}\|^2 < 10^{-4}$ for the stopping criterion. The performances of our algorithm are divided into four cases.

Case I: Comparing the proposed algorithm with various parameters θ^k are shown when we choose $\delta^k = 0.3$, $r^k = 0.2$, and $\alpha^k = \frac{k}{k+1}$. The following is a presentation of the results:

Table 1: Numerical results of different parameters θ^k.

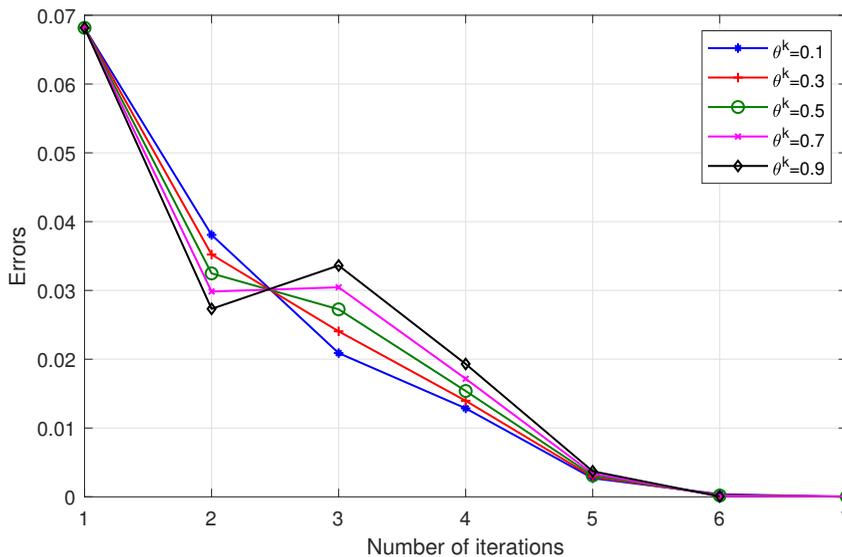| θ^k	0.1	0.3	0.5	0.7	0.9
No. of Iter.	7	7	7	7	6
CPU time(s)	6.4786	3.8913	4.1616	3.9251	3.3462

Figure 1: The Cauchy error plotting of Algorithm 3.0.1 for different parameters θ^k.

Case II: Comparing the proposed algorithm with various parameters δ^k are shown when we choose $\theta^k = 0.9$, $r^k = 0.2$, and $\alpha^k = \frac{k}{k+1}$. The following is a presentation of the results:

Table 2: Numerical results of different parameters δ^k.

δ^k	0.1	0.3	0.5	0.7	0.9
No. of Iter.	6	6	10	10	10
CPU time(s)	4.2234	0.4284	5.6592	5.5733	5.5322
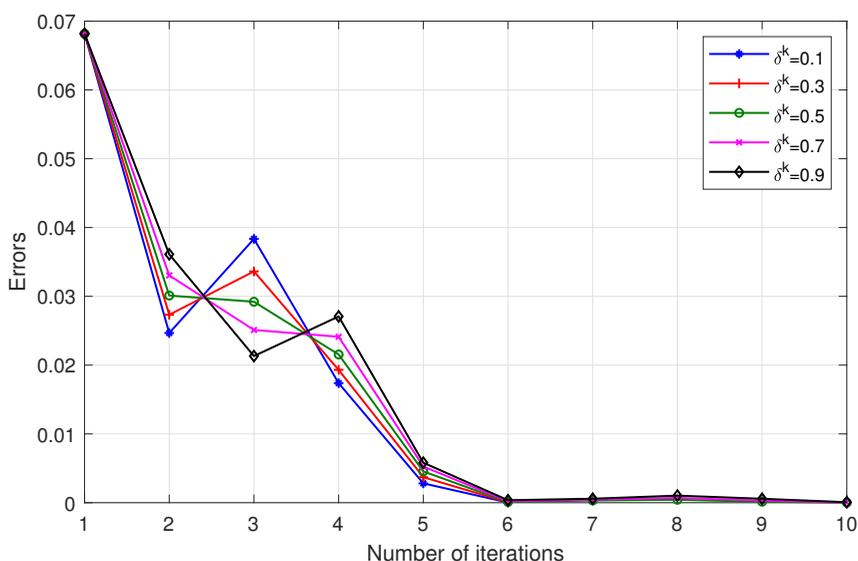
Figure 2: The Cauchy error plotting of Algorithm 3.0.1 for different parameters δ^k.

Case III: Comparing the proposed algorithm with various parameters r^k are shown when we choose $\theta^k = 0.9$, $\delta^k = 0.3$, and $\alpha^k = \frac{k}{k+1}$. The following is a presentation of the results:

Table 3: Numerical results of different parameters r^k.

r^k	0.1	0.2	0.4	0.6	0.8
No. of Iter.	8	6	6	6	5
CPU time(s)	0.5937	0.3546	0.3323	0.2833	0.2633
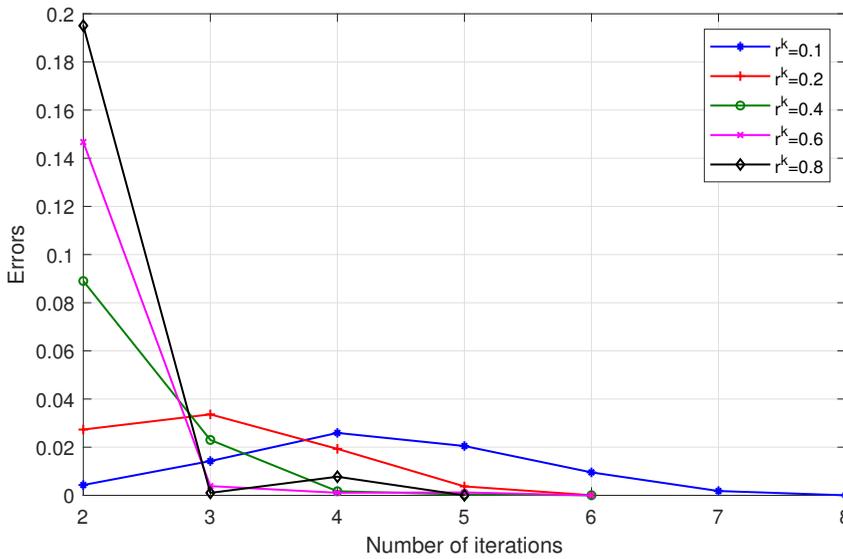
Figure 3: The Cauchy error plotting of Algorithm 3.0.1 for different parameters r^k.

Case IV: Comparing the proposed algorithm with various parameters α^k are shown by choosing $\theta^k = 0.9$, $\delta^k = 0.3$, and $r^k = 0.8$. The following is a presentation of the results:

Table 4: Numerical results of different parameters α^k.

α^k	$\frac{k}{k+1}$	$\frac{k}{2k+1}$	$\frac{k}{5k+1}$	$\frac{k}{10k+1}$	$\frac{k}{100k+1}$
No. of Iter.	5	8	11	16	2
CPU time(s)	0.2844	5.6785	7.3718	10.5273	0.9668
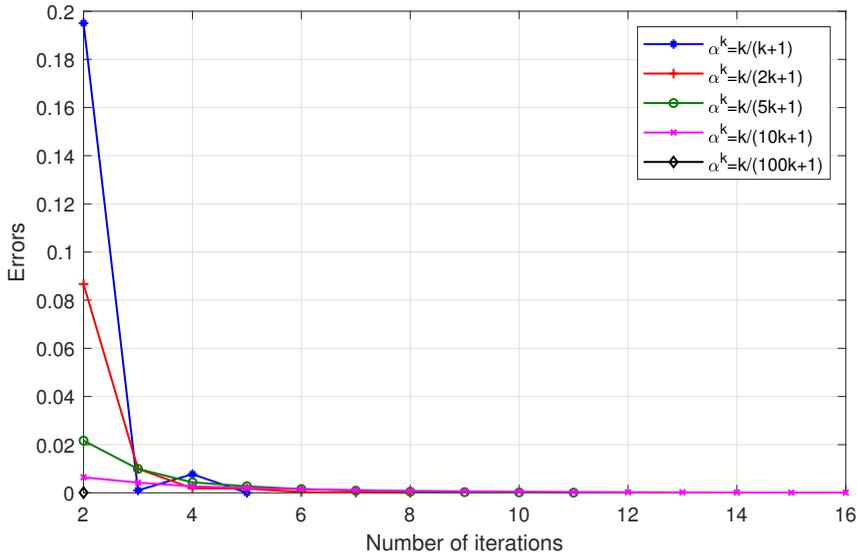
Figure 4: The Cauchy error plotting of Algorithm 3.0.1 for different parameters α^k.

From Tables 1-4 and Figures 1-4, it is evident that in all four cases mentioned, selecting the parameters $\theta^k = 0.9$, $\delta^k = 0.3$, $r^k = 0.8$, and $\alpha^k = \frac{k}{100k+1}$ provides the most favourable results when initializing with $x^{-1} = \frac{t+\sin(t)}{2}$, $x^0 = \frac{\sin(t)}{2}$, and $x^1 = \sin(t)$.

4. APPLICATION

The WHO South-East Asia Region demonstrates a steadfast commitment to enhancing healthcare services, mainly focusing on preventing, treating, and managing strokes, alongside providing excellent rehabilitative care tailored to individuals coping with stroke-related disabilities. On a global scale, stroke ranks as the second leading cause of death and the third most significant contributor to the burden of disability. Notably, one in four people worldwide faces the potential risk of experiencing a stroke during their lifetime.

Lifestyle factors contributing to stroke susceptibility include conditions such as obesity, sedentary lifestyles, tobacco use, and excessive alcohol consumption. On the medical front, risk factors include elevated blood pressure, high cholesterol levels, diabetes, and a complex interplay of personal or familial factors, including prior strokes or myocardial infarctions.

A concerning 70% of stroke cases concentrate within low- and middle-income countries, which consequently account for 87% of the total count of stroke-induced fatalities and years of life constrained by disability. To address this disparity, the WHO actively supports all constituent nations within the regional scope. Central to their mission is assisting these nations in identifying and implementing prudent "best buy" interventions, effectively reducing stroke vulnerability and broadening the availability and accessibility of exemplary stroke-focused services [26].

The investigation utilized a dataset to predict stroke, which was made available via the Kaggle website [20]. This dataset comprises a structured arrangement of 5110 rows and 12 columns. The "stroke" column is particularly significant, functioning as the dependent variable, with binary values of 1 or 0. A value of 1 signifies the presence of an associated

stroke risk, while a value of 0 indicates the absence of such a risk. In the "stroke" column, 249 rows hold 1, while 4861 rows hold 0. To improve methodology, meticulous data preprocessing balances the dataset. A comprehensive analysis is presented in Table 1.

Data preprocessing plays a crucial role in the initial stages of model building as it helps remove unwanted missing and outliers from the dataset. This process aims to improve the quality of training outcomes by addressing factors that could hinder the model's efficiency. Once a suitable dataset is collected, the next step involves data cleaning to ensure its suitability for model construction. Initially, we exclude the id column as it has minimal relevance to the model building process. Subsequently, we carefully examine the dataset to identify and correct any null values encountered. In this case, the column bmi has null values filled with the mean of the column data. We remove null values from the data set before applying the data to classify the data.

Table 1: Stroke dataset attributes information.

Attribute Name	Definitions and Encoding
id	Patient ID
Input	
gender	Gender of Patient:
	1 := Male
	2 := FeMale
	3 := Other
age	Age of Patient
hypertension	0 := no hypertension
	1 := suffering from hypertension
heart_disease	0 := no heart disease
	1 := suffering from heart disease
ever_married	1:=Yes
	2:=No
work_type	0:=children
	1 := Govt_jov
	2 := Never_worked
	3 := Private
	4 := Self-employed
Residence_type	1:=Rural
	2 := Urban
avg_glucose_level	average glucose level in blood
bmi	body mass index
smoking_status	0 := formerly smoked
	1 := never smoked
	2 := smokes
	3 := Unknown
Output	
stroke	0 := no stroke
	1 := suffered stroke

In this study, our focus is on extreme learning machine (ELM) [14] applied to single-hidden layer feedforward neural networks (SLFNs), which are defined as follows:

Let $x^n = [x^1, x^2, ..., x^N]^T \in \mathcal{R}^n$ is an input training data and $t^n = [t^1, t^2, ..., t^N]^T \in \mathcal{R}^m$ is a target, standard SLFNs with M hidden nodes. Assume that activation function $\mathcal{A}$ is

$$O^j = \sum_{i=1}^{M} \varphi^i \mathcal{A}(\langle \omega^i, x^j \rangle + b^i)$$

where φ^i is the optimal output weight at the i-th hidden node, ω^i is weight, and b^i is bias. The hidden layer output matrix $\mathcal{H}$ is generated as follows:

$$\mathcal{H} = \begin{bmatrix} \mathcal{A}(\langle \omega^1, x^1 \rangle + b^1) & \cdots & \mathcal{A}(\langle \omega^M, x^1 \rangle + b^M) \\ \vdots & \ddots & \vdots \\ \mathcal{A}(\langle \omega^1, x^N \rangle + b^1) & \cdots & \mathcal{A}(\langle \omega^M, x^N \rangle + b^M) \end{bmatrix}$$

The main goal of ELM is to determine the optimal output weight vector, denoted as $\varphi = [\varphi^{1T}, \varphi^{2T}, ..., \varphi^{MT}]^T$, that satisfies the equation $\mathcal{H}\varphi = \mathcal{T}$. Here, $\mathcal{T} = [t^{1T}, t^{2T}, ..., t^{NT}]^T$ represents the training target data. However, a challenge arises in obtaining the solution

$\varphi = \mathcal{H}^{\ddagger}\mathcal{T}$, where $\mathcal{H}^{\ddagger}$ denotes the Moore-Penrose generalized inverse of $\mathcal{H}$, can be challenging if the matrix $\mathcal{H}$ does not exist. Consequently, resolving this issue through convex minimization allows us to find a solution for φ.

In the section, we will thoroughly review pertinent experimental trials for our classification problem. We can apply Algorithm 3.0.1 by considering it within two problem models, as follows:

(i) The regularization of least square problem (RLSP): for $\lambda > 0$

$$(4.11) \qquad \min_{\varphi \in \mathbb{R}^M} \left\{ \frac{1}{2} \|\mathcal{H}\varphi - \mathcal{T}\|_2^2 + \lambda \|\varphi\|_1 \right\}.$$

The problem in equation (4.11), known as the least absolute shrinkage and selection operator (LASSO), was introduced by Tibshirani [28]. Our Algorithm 3.0.1 addresses this problem with the setting operators: $\mathsf{F}(\varphi) \equiv \nabla(\frac{1}{2}\|\mathcal{H}\varphi - \mathcal{T}\|_2^2)$ and $\mathsf{G}(\varphi) \equiv \partial(\lambda\|\varphi\|_1)$.

(ii) The regularization of least square problem with constrained by closed convex set (RLSPC): for $\lambda, \mu > 0$

$$(4.12) \qquad \min_{\varphi \in E} \left\{ \frac{1}{2} \|\mathcal{H}\varphi - \mathcal{T}\|_2^2 + \lambda \|\varphi\|_1 \right\}.$$

where $E = \{\varphi : \|\varphi\|_2^2 \leq \mu\}$. Our Algorithm 3.0.1 addresses this problem with the setting operators: $\mathsf{F}(\varphi) \equiv \nabla(\frac{1}{2}\|\mathcal{H}\varphi - \mathcal{T}\|_2^2)$ and $\mathsf{G}(\varphi) \equiv \partial(\lambda\|\varphi\|_1)$.

The study emphasises assessing crucial performance metrics: accuracy, precision, recall, and F1-score [29]. These metrics are determined by considering specific values, such as true negatives ($\mathsf{T_N}$), false positives ($\mathsf{F_P}$), false negatives ($\mathsf{F_N}$), and true positives ($\mathsf{T_P}$).

$$Accuracy = \frac{\mathsf{T_P} + \mathsf{T_N}}{\mathsf{T_P} + \mathsf{T_N} + \mathsf{F_P} + \mathsf{F_N}} \times 100\%;$$

$$Precision = \frac{\mathsf{T_P}}{\mathsf{T_P} + \mathsf{F_P}} \times 100\%;$$

$$Recall = \frac{\mathsf{T_P}}{\mathsf{T_P} + \mathsf{F_N}} \times 100\%;$$

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

The standard binary cross entropy loss function [13] is the cross entropy loss when only two classes are involved. We can calculate the loss by calculating the following mean:

$$Loss = -\frac{1}{n} \sum_{i=1}^{n} \gamma_i \log \hat{\gamma}_i + (1 - \gamma_i) \log(1 - \hat{\gamma}_i),$$

where n is the number of training examples, γ_i is target label for training example i-th, and $\hat{\gamma}_i$ is model with neural network weights i-th.

This paper is focused on the exclusive use of the training dataset. Before constructing the model, the dataset underwent a trisection, creating two distinct subsets: a training set encompassing 70% of the dataset and a complementary test set containing the remaining 30%. The computational process commences with configuring the activation function as sigmoid, concurrently involving the implementation of $\mathsf{M} = 150$ hidden nodes. The stopping criteria is the number of iteration 900. After this, a performance assessment is conducted to compare the effectiveness of the algorithm we studied with the other algorithms mentioned above. This assessment is conducted meticulously, adhering to the parameters in Table 2.

Table 2: Chosen parameters of each algorithm.

Algorithm	r^k	θ^k	δ^k	α^k	λ	μ
IPA	$\frac{1.9999}{\|\mathcal{H}\|^2}$	$\frac{1}{\|x^k-x^{k-1}\|^5+k^5}$	-	-	10^{-5}	-
SIPA	$\frac{1.9999}{\|\mathcal{H}\|^2}$	$\frac{1}{\|x^k-x^{k-1}\|^6+k^6}$	-	-	10^{-5}	-
TSIPA	$\frac{1.9999}{\|\mathcal{H}\|^2}$	$\frac{1}{\|x^k-x^{k-1}\|^3+k^3}$	$\frac{1}{\|x^k-x^{k-1}\|^3+k^3}$	1	10^{-5}	-
Algorithm 3.0.1(RLSP)	$\frac{0.3}{\|\mathcal{H}\|^2}$	$\frac{2^{15}}{\|x^k-x^{k-1}\|^3+k^3+2^{15}}$	$\frac{2^{15}}{\|x^k-x^{k-1}\|^3+k^3+2^{15}}$	$\frac{0.9k}{k+1}$	10^{-5}	-
Algorithm 3.0.1(RLSPC)	$\frac{0.5}{\|\mathcal{H}\|^2}$	$\frac{2^{15}}{\|x^k-x^{k-1}\|^3+k^3+2^{15}}$	$\frac{2^{15}}{\|x^k-x^{k-1}\|^3+k^3+2^{15}}$	$\frac{0.9k}{k+1}$	10^{-5}	9
Algorithm 3.0.1-(i)(RLSPC)	$\frac{1.9999}{\|\mathcal{H}\|^2}$	$\frac{1}{\|x^k-x^{k-1}\|^5+k^5}$	0	0.9999	10^{-5}	4
Algorithm 3.0.1-(ii)(RLSC)	$\frac{1.9999}{\|\mathcal{H}\|^2}$	0	$\frac{1}{\|x^k-x^{k-1}\|^6+k^6}$	0.9999	10^{-5}	4
Algorithm 3.0.1-(iii)(RLSPC)	$\frac{1.9999}{\|\mathcal{H}\|^2}$	0	0	0.9999	10^{-5}	25
Algorithm 3.0.1-(iv)(RLSPC)	$\frac{0.3}{\|\mathcal{H}\|^2}$	$\frac{2^{12}}{\|x^k-x^{k-1}\|^3+k^3+2^{12}}$	$\frac{2^{15}}{\|x^k-x^{k-1}\|^3+k^3+2^{15}}$	1	10^{-5}	9
Algorithm 3.0.1-(v)(RLSPC)	$\frac{1.9999}{\|\mathcal{H}\|^2}$	$\frac{1}{\|x^k-x^{k-1}\|^5+k^5}$	0	1	10^{-4}	4
Algorithm 3.0.1-(vi)(RLSPC)	$\frac{1.9999}{\|\mathcal{H}\|^2}$	0	$\frac{1}{\|x^k-x^{k-1}\|^6+k^6}$	1	10^{-4}	9

Table 3: The performance of each algorithm.

| Algorithm	Training Time	Precision	Recall	F1-score	Accuracy
IPA	0.0206	96.37	97.94	97.15	94.50
SIPA	0.0193	95.90	99.50	97.67	95.45
TSIPA	0.0235	98.06	82.34	89.51	81.53
Algorithm 3.0.1(RLSP)	0.0271	95.85	99.93	97.85	95.79
Algorithm 3.0.1(RLSPC)	0.0196	95.85	99.93	97.85	95.79
Algorithm 3.0.1-(i)(RLSPC)	0.0197	95.72	99.93	97.78	95.66
Algorithm 3.0.1-(ii)(RLSPC)	0.0192	95.72	100.00	97.81	95.72
Algorithm 3.0.1-(iii)(RLSPC)	0.0206	95.72	100.00	97.81	95.72
Algorithm 3.0.1-(iv)(RLSPC)	0.0206	95.72	99.93	97.78	95.66
Algorithm 3.0.1-(v)(RLSPC)	0.1375	95.72	100.00	97.81	95.72
Algorithm 3.0.1-(vi)(RLSPC)	0.0189	95.72	100.00	97.81	95.72

Table 3 shows that the algorithms we study have the highest efficiency in accuracy, recall, and F1-score. It has the highest probability of correctly classifying stroke compared to the mentioned algorithms.

The performance results of the methods used in the literature with the same data set. This study was compared to the literature [24] in terms of accuracy, precision, recall, and F1-score. The literature [24] used machine learning algorithms such as Logistic Regression, Decision Tree Classification, Random Forest Classification, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes Classification. A meticulous dataset division into discrete training and testing subsets was enacted in data preprocessing. This division entailed an allocation of 80% for training data and an accompanying 20% for testing data. The resultant outcomes of these rigorous analytical endeavours are systematically showcased in Table 4.

Table 4: The performance of ML algorithms.

Algorithm	Precision	Recall	F1-score	Accuracy
Logistic Regression	77.50	77.60	77.60	78.00
Decision Tree Classification	77.50	77.50	77.60	66.00
Random Forest Classification	72.00	73.50	72.70	73.00
K-Nearest Neighbors (KNN)	77.40	83.70	80.40	80.00
Support Vector Machine (SVM)	78.60	83.80	81.10	80.00
Naive Bayes Classification	79.20	85.70	82.30	82.00
Algorithm 3.0.1(RLSP)	**95.85**	**99.93**	**97.85**	**95.79**
Algorithm 3.0.1(RLSPC)	**95.85**	**99.93**	**97.85**	**95.79**

Table 4 shows that of all the algorithms chosen on the stroke dataset. The Algorithm 3.0.1(RLSP) and Algorithm 3.0.1(RLSPC) performs best with an accuracy of 95.79%. Our Algorithm 3.0.1(RLSP) and Algorithm 3.0.1(RLSPC) also achieved the highest precision, recall, and F1-score compared to the literature [24].
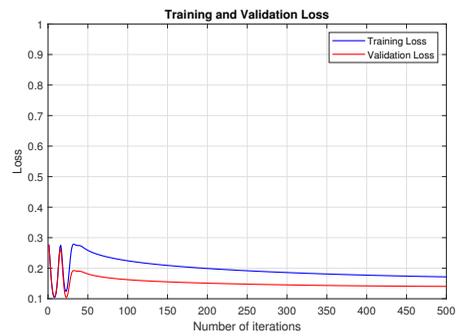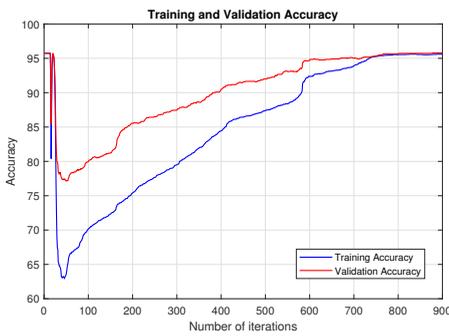
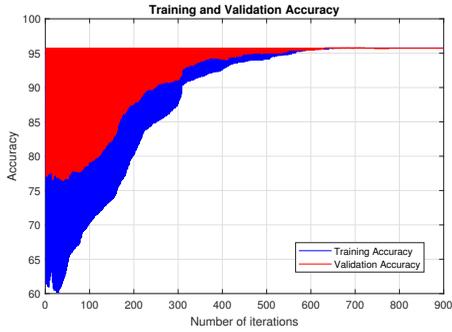**Figures 5-6:** Accuracy and Loss graph of the Algorithm 3.0.1(RLSP).
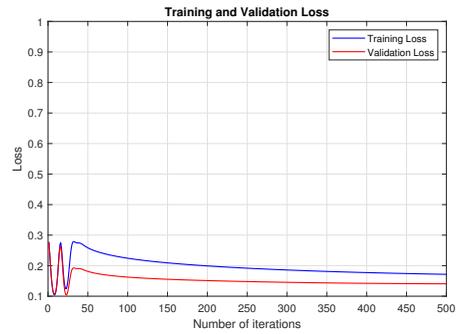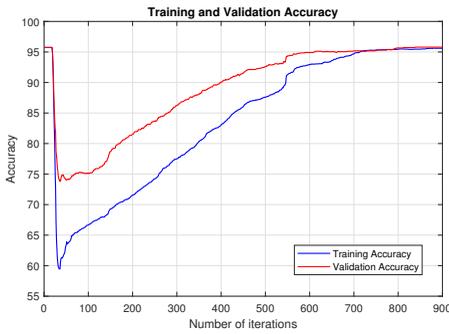
Figures 7-8: Accuracy and Loss graph of the Algorithm 3.0.1(RLSPC).
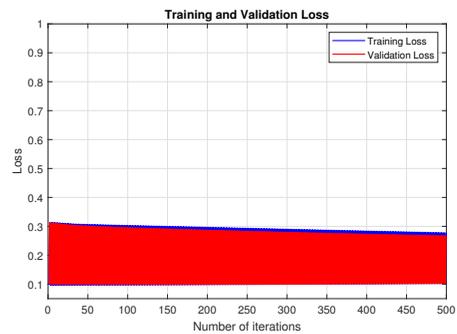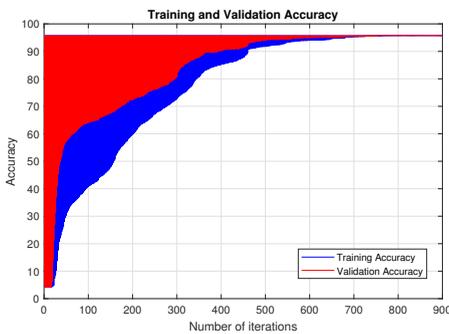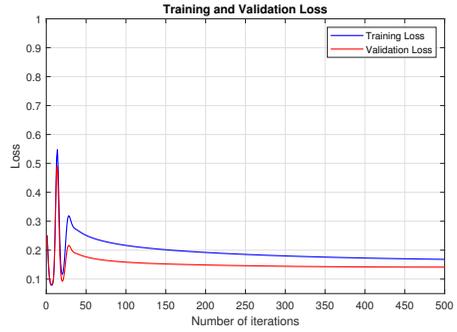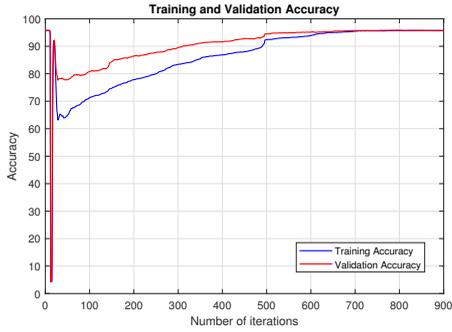
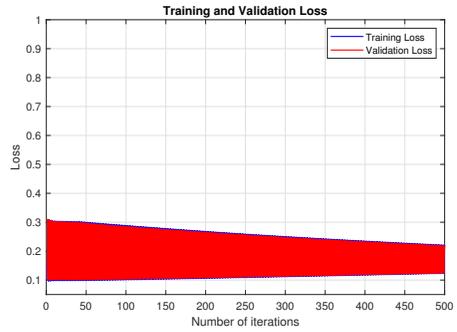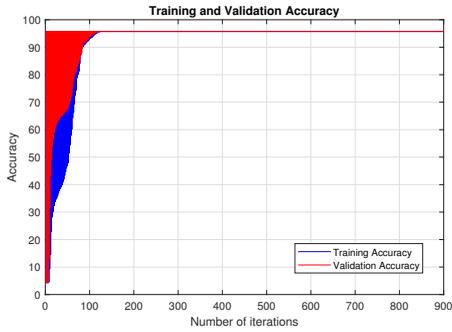**Figures 9-10:** Accuracy and Loss graph of the Algorithm 3.0.1-(i)(RLSPC).
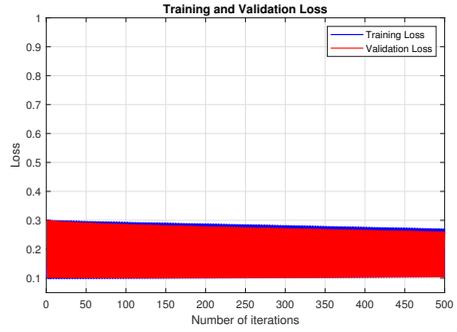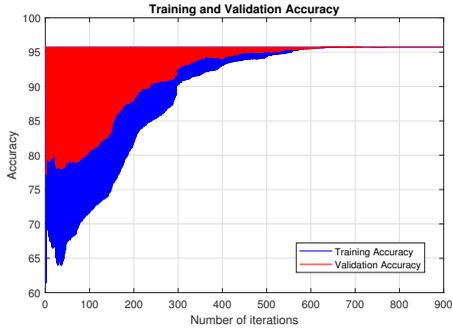
Figures 11-12: Accuracy and Loss graph of the Algorithm 3.0.1-(ii)(RLSPC).
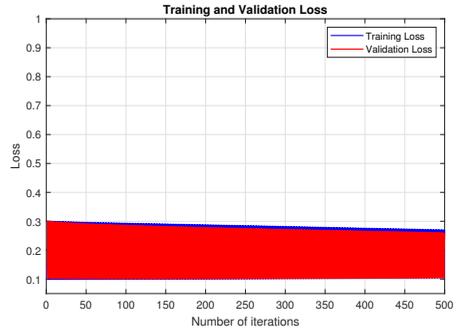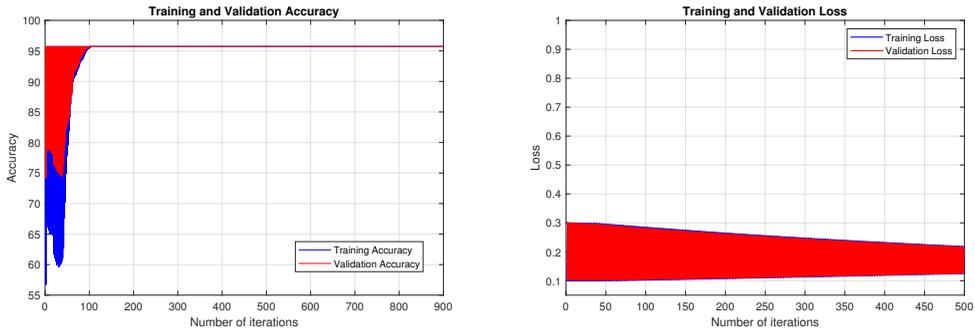

Figures 13-14: Accuracy and Loss graph of the Algorithm 3.0.1-(iii)(RLSPC).

Figures 15-16: Accuracy and Loss graph of the Algorithm 3.0.1-(iv)(RLSPC).

Figures 17-18: Accuracy and Loss graph of the Algorithm 3.0.1-(v)(RLSPC).

Figures 19-20: Accuracy and Loss graph of the Algorithm 3.0.1-(vi)(RLSPC).

From Figures 5-20, we can see that Training Loss and Validation Loss consistently decrease until it will remain constant at some point. The two graphs have very little gaps between them. The recognition that the algorithms we study function as an appropriately good fitting model implies its adeptness in effectively acquiring knowledge from the training dataset.

5. CONCLUSIONS

We have suggested a two inertial technique with a projection Mann forward-backward splitting algorithm for finding solutions of variational inclusion problems in Hilbert spaces. Weak convergence result for the proposed method was established under some mild conditions. We provide a numerical example in infinitely dimensional spaces to support the main result. Moreover, we applied our findings to predict occurrences of stroke. The creation of a machine learning algorithm has the potential to enable the early anticipation of strokes, thus mitigating their potentially significant consequences. Our algorithm is based on the fundamental principles of the extreme learning machine model, which has been carefully tailored to excel in classification tasks. Finally, we show the performance of our algorithm by comparing it with the other three algorithms and the literature mentioned. These results show our algorithm is better than the three algorithms and literature.

Data Availability

The dataset used in this research is publicly available at the Kaggle on
https://www.kaggle.com/fedesoriano/stroke-prediction-dataset

Conflicts of Interest

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

Author Contributions

Writing this paper, P.K.; software, P.P.; review and editing paper, W.C.; review and editing software, D.Y. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] Abubakar, J.; Kumam, P.; Garba, A. I.; Abdullahi, M. S.; Ibrahim, A. H.; Sitthithakerngkiet, K. An inertial iterative scheme for solving variational inclusion with application to Nash-Cournot equilibrium and image restoration problems. *Carpathian J. Math.* **37** (2021), no. 3, 361-380.

[2] Agarwal, R. P.; Huang, N. D.; Tan, M. Y. Sensitivity analysis for a new system of generalized nonlinear mixed quasi-variational inclusions. *Appl. Math. Lett.* **17** (2004), no. 3, 345-352.

[3] Ansari, Q. H.; Konnov, I. V.; Yao, J. C. Existence of a solution and variational principles for vector equilibrium problems. *J. Optim. Theory Appl.* **110** (2001), no. 3, 481-492.

[4] Alvarez, F.; Attouch, H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **9** (2001), 3-11.

[5] Auslender, A.; Teboulle, M.; Ben-Tiba, S. A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim.: A Tribute to Olvi Mangasarian* **1** (1999), 31-40.

[6] Bauschke, H. H.; Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics. Springer, New York.

[7] Bianchi, M.; Schaible, S. Generalized monotone bifunctions and equilibrium problems. *J. Optim. Theory Appl.* **90** (1996), no. 1, 31-43.

[8] Blum, E. From optimization and variational inequalities to equilibrium problems. *Math. Stud.* **63** (1994), 123-145.

[9] Chang, S. S. Set-valued variational inclusions in Banach spaces. *J. Math. Anal. Appl.* **248** (2000), no. 2, 438-454.

[10] Chang, S. S.; Kim, J. K.; Kim, K. H. On the existence and iterative approximation problems of solutions for set-valued variational inclusions in Banach spaces. *J. Math. Anal. Appl.* **268** (2002), no. 1, 89-108.

[11] Fu, J. Y. Generalized vector quasi-equilibrium problems. *Math. Methods Oper. Res.* **52** (2000), no. 1.

[12] Goebel, K.; Kirk, W. A. *Topics in metric fixed point theory.* Cambridge university press, 1990.

[13] Ho, Y.; Wookey, S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access* **8** (2019), 4806-4813.

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

[15] Iyiola, O. S.; Shehu, Y. Convergence results of two-step inertial proximal point algorithm. *Appl. Numer. Math.* **182** (2022), 57-75.

[16] López, G.; Martín-Márquez, V.; Wang, F.; Xu, H. K. Forward-backward splitting methods for accretive operators in Banach spaces. *Abstr. Appl. Anal.* **2012** (2012).

[17] Lin, L. J.; Ansari, Q. H.; Wu, J. Y. Geometric properties and coincidence theorems with applications to generalized vector equilibrium problems. *J. Optim. Theory Appl.* **117** (2003), 121-137.

[18] Lorenz, D. A.; Pock, T. An inertial forward-backward algorithm for monotone inclusions. *J. Math. Imaging Vis.* **51** (2015), 311-325.

[19] Luc, D. T.; Tan, N. X. Existence conditions in variational inclusions with constraints. *Optim.* **53** (2004), no. 5-6, 505-515.

[20] Dataset named 'Stroke Prediction Dataset' from Kaggle: https://www.kaggle.com/fedesoriano/stroke-prediction-dataset.

[21] Peeyada, P.; Suparatulatorn, R.; Cholamjiak, W. An inertial Mann forward-backward splitting algorithm of variational inclusion problems and its applications. *Chaos Solit. Fractals* **158** (2022), 112048.

[22] Marino, G.; Xu, H. K. Convergence of generalized proximal point algorithms. *Commun. Pure Appl. Anal.* **3** (2004), no. 4, 791-808.

[23] Moudafi, A.; Oliny, M. Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.* **155** (2003), no. 2, 447-454.

[24] Sailasya, G.; Kumari, G. L. A. Analyzing the performance of stroke prediction using ML classification algorithms. *Int. J. Adv. Comput. Sci. Appl.* **12** (2021), no. 6.

[25] Seangwattana, T.; Sombut, K.; Arunchai, A.; Sitthithakerngkiet, K. A modified Tseng's method for solving the modified variational inclusion problems and its Applications. *Symmetry* **13** (2021), no. 12, 2250.

[26] Singh, P. K., (2021). World Stroke Day. WHO Regional Director for South-East Asia: https://www.who.int/southeastasia/news/detail/28-10-2021-world-stroke-day#: :text=Globally%2C%20stroke%20is%20the%20second,tobacco%20use%20and %20alcohol%20abuse.

[27] Tuan, L.A.; Sach, P.H. Existence of solutions of generalized quasivariational inequalities with set-valued maps. *Acta Math. Vietnam.* **29** (2004), 309.

[28] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol* **58** (1996), 267-288.

[29] Wardhani, N. W. S.; Rochayani, M. Y.; Iriany, A.; Sulistyono, A. D.; Lestantyo, P. Cross-validation metrics for evaluating classification performance on imbalanced data. In 2019 international conference on computer, control, informatics and its applications (IC3INA). *IEEE*. (2019), 14-18.

SCHOOL OF SCIENCE
UNIVERSITY OF PHAYAO
PHAYAO 56000, THAILAND
Email address: `pronpat.pee@gmail.com`
Email address: `pronkamonkitisak@gmail.com`
Email address: `watcharaporn.ch@up.ac.th`
Email address: `damrongsak.ya@up.ac.th`