AIMS Mathematics, 11(6): 18943–18969
DOI:10.3934/math.2026771
Received: 03 May 2026
Revised: 10 June 2026
Accepted: 22 June 2026
Published: 26 June 2026

https://www.aimspress.com/journal/Math

Research article

A new Dai-Liao-type algorithm for efficient neural network learning in medical diagnosis

Mehamdia Abd Elhamid[1], Raouf Ziadi[2,*], Alaa Luqman Ibrahim[3], Mohammed A. Saleh[4] and Abdulgader Z. Almaymuni[4,*]

[1] University M'Hamed Bougara of Boumerdes, Boumerdes 35000, Algeria

[2] Laboratory of Fundamental and Numerical Mathematics (LMFN), Department of Mathematics, University Setif-1-Ferhat Abbas, Setif, Algeria

[3] Department of Mathematics, College of Science, University of Zakho, Zakho, Kurdistane Region, Iraq

[4] Department of Cybersecurity, College of Computer, Qassim University, Saudi Arabia

* **Correspondence:** Email: ziadi.raouf@gmail.com, almaymuni@qu.edu.sa.

Abstract: Conjugate gradient (CG) methods are considered among the most efficient methods for solving optimization problems thanks to their straightforward iterative process and low memory requirements. In the present work, we propose a combined CG method to address large-scale problems, with a particular application to training artificial neural networks (ANNs) for early breast cancer prediction and electrocardiogram (ECG) classification. Under the strong Wolfe line search conditions, the global convergence was demonstrated under mild assumptions and the generated descent direction and the convergence features of the suggested approach are examined. The proposed approach was successfully applied to train neural networks for early breast cancer prediction, achieving an accuracy of 98.24%, with precision, recall, and F1-score values of 0.99, 0.97, and 0.98, respectively. It also reduces the final mean squared error by over 52% and exhibited faster convergence with smoother training dynamics. Furthermore, on the ECG classification dataset, the proposed hybrid Dai-Liao (hDL^{+}) achieves an accuracy of 80.45%, demonstrating strong generalization performance across different medical diagnostic applications. Comparisons with recent CG methods on a set of test problems from the CUTE library confirmed the robustness and efficiency of the proposed method.

Keywords: optimization method; Dai-Liao method; artificial neural networks; breast cancer prediction; ECG classification; medical decision support systems
Mathematics Subject Classification: 65K05, 90C30, 90C56

1. Introduction

Breast cancer is the most common diagnosed cancer in women worldwide and continues to be a major public health problem. Recent global cancer statistics show that in 2020, there were more than 2.3 million new breast cancer cases reported worldwide and about 685,000 deaths related to this disease [17, 18]. It is one of the leading causes of cancer death in women, especially in low- and middle-income countries where facilities for early screening and diagnosis are often unavailable. For this reason, early detection is important to increase patient survival and the number of treatment options since the diagnosis can be made in the initial stages with less aggressive therapies and a much better prognosis [5].

In clinical practice, breast cancer diagnosis is mainly based on medical imaging modalities such as mammography, ultrasound and magnetic resonance imaging (MRI) that are selected according to patient-specific risk factors and breast tissue characteristics [19]. Interpretation of such data, however, remains difficult and is prone to inter-observer variability.

Recently, artificial intelligence (AI) and machine learning (ML) have rapidly developed and have resulted in computer-aided diagnosis (CAD) systems that assist clinicians in detecting and classifying cancerous lesions with improved accuracy and consistency. Among these techniques, artificial neural networks (ANNs) have proved to be promising tools to model complex nonlinear relationships and to extract discriminative patterns from high-dimensional medical data [9, 45]. The overall performance of ANN-based diagnostic systems, although successful, highly depends on the efficiency, stability and convergence behavior of the underlying training algorithms.

Recent studies have also demonstrated the effectiveness of advanced mathematical modeling and optimization techniques in analyzing complex dynamic systems. For instance, Dai and Wu [14] employed sophisticated econometric models to investigate systemic risk spillovers in commodity markets under oil shocks, highlighting the importance of robust computational approaches for extracting reliable patterns from large-scale and highly interconnected data. Such findings further motivate the development of efficient optimization algorithms for data-driven applications.

Recent studies on nonlinear and fractional dynamical systems have further highlighted the importance of stability analysis and robust computational techniques in complex modeling problems [32–36]. These developments reinforce the need for stable and efficient optimization algorithms in data-driven learning frameworks.

There are a number of studies performed on the use of neural networks and hybrid learning strategies in the diagnosis of cancer. Hu et al. [22] compared supervised and unsupervised learning methods for cancer cell classification. They used a single-hidden-layer feedforward neural network trained with backpropagation for supervised learning, and fuzzy and non-fuzzy c-means clustering for unsupervised learning. Their experiments on 467 tumour images reported a classification accuracy of up to 96.9% using neural networks. Won and Cho [41] suggested an ensemble neural network classifier with negatively correlated features for cancer classification based on gene expression data. Their method achieved better recognition performance than the state-of-the-art on several benchmark data sets. Similarly, Xu et al. [42] proposed a hybrid model that combines probabilistic neural networks with a discrete binary particle swarm optimisation algorithm for gene selection and dimensionality reduction, which achieved promising classification accuracy on lymphoma datasets. Bevilacqua et al. [4] focused on the optimisation of ANN topology by using a multi-objective genetic

algorithm applied to the Wisconsin Breast Cancer Database (WBCD). Their results showed the importance of network architecture design and optimisation strategies for improving the classification accuracy.

The results of these studies confirm the effectiveness of neural networks for cancer diagnosis, but they also highlight the need for more robust and faster optimisation techniques to improve training efficiency and generalization performance.

In addition to breast cancer diagnosis, cardiovascular diseases (CVDs) represent another major medical challenge due to their high incidence and mortality rates, particularly among aging populations. Electrocardiogram (ECG) signals play a fundamental role in the early detection and monitoring of cardiac abnormalities. Recent advances in deep neural networks (DNNs) have significantly improved the performance of ECG-based diagnostic systems by enabling automatic feature extraction from complex temporal and nonlinear signal patterns [27, 37, 40]. These methods have demonstrated strong capability in accurately classifying different types of arrhythmias and cardiovascular conditions using standard benchmark datasets such as the MIT-BIH Arrhythmia Database.

Inspired by these studies, we propose the hybrid Dai-Liao method (hDL$^+$), a novel CG-type approach for artificial neural network training. The proposed method is designed to enhance numerical stability, convergence speed, and memory efficiency. Its effectiveness is evaluated on two benchmark medical applications, namely breast cancer classification using the Wisconsin WBCD dataset and ECG signal classification for CVD detection, demonstrating its robustness and generalization capability across different medical domains.

This work is organized as follows. In Section 2, the neural network training framework for breast cancer prediction is introduced. In Section 3, related CG methods are reviewed. In Section 4, the proposed hDL$^+$ algorithm is introduced, and the sufficient descent condition is proved. In Section 5, the global convergence of the proposed method with the strong Wolfe line search is proved. The numerical results are contained in Section 6. Finally, a summary of the paper is made in Section 7.

2. Neural network training for early breast cancer prediction

This section presents the application of the proposed CG algorithm to train ANNs for early breast cancer diagnosis. We detail the background, dataset, network architecture, and performance metrics used in this study.

2.1. Neural network training formulation

Training an ANN can be formulated as an unconstrained optimization problem, where the objective is to minimize a loss function that quantifies the discrepancy between predicted outputs and ground-truth labels. Let $\mathbf{w} \in \mathbb{R}^n$ denote the vector of all network weights and biases. In this work, the mean squared error (MSE) function is employed and defined as

$$E(\mathbf{w}) = \sum_{p=1}^{P} \sum_{j=1}^{N_L} \left(y_{j,p}^L - t_{j,p}\right)^2, \tag{2.1}$$

where P represents the number of training samples, N_L denotes the number of neurons in the output layer, $t_{j,p}$ is the target output of neuron j for the p-th sample, and $y_{j,p}^L$ is the corresponding network

output.

CG methods are particularly well suited for large-scale neural network training due to their fast convergence properties and minimal memory requirements [23, 24]. In this study, the proposed hDL$^+$ algorithm is employed as the core optimization strategy, and its performance is benchmarked against the standard Hestenes-Stiefel (HS) method.

2.2. Dataset description

The Breast Cancer Wisconsin (Diagnostic) dataset, which was acquired from Kaggle*, was used for experimental validation. Each of the 569 patient samples in the dataset, 212 malignant and 357 benign cases, is characterized by 30 real-valued features that were taken from digitized fine-needle aspirate (FNA) biopsy images.

Numerous characteristics of cell nuclei, including radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension, are described by these features. To enhance numerical stability and convergence behavior, all characteristics were normalized before training. Prior to training, the diagnosis labels were converted into binary values, where malignant cases were encoded as 1 and benign cases as 0. The input features were normalized using MATLAB default preprocessing functions to improve numerical stability and facilitate efficient network training. The dataset was randomly divided into training (70%), validation (15%), and testing (15%) subsets.

In addition to the breast cancer dataset, the proposed approach was further evaluated using the MIT-BIH Arrhythmia ECG dataset, which is widely used for cardiovascular disease classification tasks. The dataset consists of 1019 ECG signal samples, each represented by 187 features extracted from physiological recordings. The classification problem involves five categories: Normal Beat (0), Supraventricular Ectopic Beat (1), Ventricular Ectopic Beat (2), Fusion Beat (3), and Unknown Beat (4), which reflect different types of cardiac arrhythmias.

Similar preprocessing steps were applied to the ECG dataset to ensure consistent training conditions. Due to class imbalance, an oversampling strategy was employed to balance all classes by replicating minority samples. After balancing, the dataset was randomly shuffled and normalized prior to training. The dataset was then divided into training (70%), validation (15%), and testing (15%) subsets to ensure fair evaluation and avoid bias in performance results.

2.3. Neural network architecture

Four layers make up the selected ANN architecture: an input layer, two hidden layers, and an output layer. The two hidden layers contain 10 and 5 neurons, respectively, while the output layer consists of a single neuron for binary classification (malignant or benign). Sigmoid activation functions are used in the hidden layers, whereas a sigmoid (logistic) activation function is employed in the output layer. The network architecture is illustrated in Figure 1.

*https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data

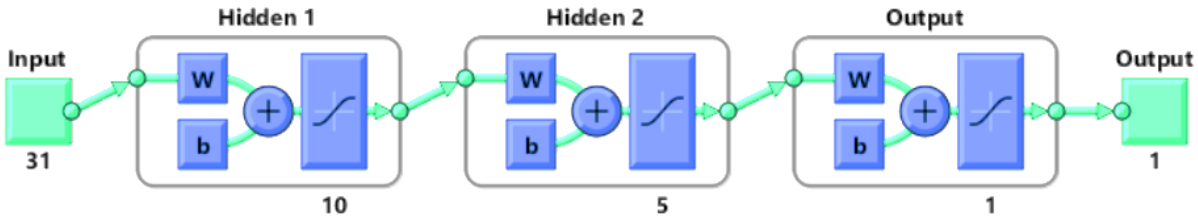

Figure 1. The architecture of a neural network to predict breast cancer early.

The neural network was implemented using MATLAB `patternnet` framework with two hidden layers containing 10 and 5 neurons, respectively. Training was performed using the proposed CG-based optimization algorithm. The maximum number of training epochs was set to 1000, and the training process was terminated when the MSE reached the predefined goal value of 10^{-5} or when the maximum number of epochs was exceeded. The network weights and biases were automatically initialized using MATLAB default random initialization procedure before training.

The same neural network architecture and training configuration were also applied to the ECG dataset to ensure a fair comparison and consistent experimental setup. In this case, the input layer consists of 187 features corresponding to the extracted ECG signal descriptors, while the output layer is modified to include 5 neurons representing the five ECG classes. All other architectural and training parameters were kept identical to those used in the breast cancer classification experiments.

2.4. Performance metrics

Let TP, TN, FP, and FN denote the elements of the confusion matrix corresponding to true positives, true negatives, false positives, and false negatives, respectively. The model performance is assessed using the following standard measures:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2.2}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2.3}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{2.4}$$

$$\text{F1-score} = 2 \cdot \frac{TP}{2TP + FP + FN}. \tag{2.5}$$

In medical classification problems, particular emphasis is placed on minimizing FN, as failing to detect positive cases may have serious clinical implications.

3. Conjugate gradient methods

Minimizing the MSE in (2.1) amounts to solving the following nonlinear minimization problem:

$$f^* = \min_{x \in \mathbb{R}^n} f(x), \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function.

Before introducing our methodology, we briefly review some facts about CG methods for the reader's convenience. These have emerged as powerful tools for solving large-scale minimization problems, playing a fundamental role in various scientific and engineering applications. Their appeal lies in their ability to generate decent directions with minimal storage requirements and fast convergence properties, making them particularly suitable for training machine learning models and neural networks. Starting from a point $x_0 \in \mathbb{R}^n$, the sequence of points $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ is generated by the following recursive scheme:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k \in \mathbb{N}, \tag{3.1}$$

where α_k is the step size that is usually determined using inexact line searches, which guarantee the steps are neither too long nor too short, while d_k is the search direction, typically computed iteratively:

$$d_{k+1} = -H_{k+1} + \beta_k d_k;\ d_0 = -H_0, \tag{3.2}$$

with $H_k = \nabla f(x_k)$, and β_k is the conjugate parameter.

The literature has presented a number of classical CG methods, such as the Fletcher-Reeves (FR) method [16], the Polak-Ribire-Polyak (PRP) method [38, 39], the Hestenes-Stiefel (HS) method [20], and the Dai-Yuan (DY) method [10]. However, these classical methods suffer from certain limitations; for instance, the FR method may exhibit jamming behavior, while the PRP method does not guarantee global convergence for nonconvex functions without additional conditions.

Over the past decades, numerous variants and modifications of CG methods have been proposed to enhance their numerical stability, global convergence properties, and practical performance [29, 31]. Among these, hybrid approaches that combine the strengths of different classical formulations have attracted considerable attention due to their improved efficiency and robustness [7, 8].

In the context of developing this method, Mehamdia et al. [30] proposed two new efficient CG coefficients, denoted by β_k^{IDY} and β_k^{IFR}, whose corresponding conjugate parameters are

$$\beta_k^{IDY} = \frac{||H_{k+1}||^2 - \frac{\eta_1 \left|H_{k+1}^T (H_{k+1} - H_k)\right| . \left|H_{k+1}^T d_k\right|}{||H_{k+1} - H_k|| . ||d_k||}}{d_k^T\ (H_{k+1} - H_k)}, \text{ and } \beta_k^{IFR} = \frac{||H_{k+1}||^2 - \frac{\eta_2 \left|H_{k+1}^T (H_{k+1} - H_k)\right| . \left|H_{k+1}^T d_k\right|}{||H_{k+1} - H_k|| . ||d_k||}}{||H_k||^2},$$

where η_1,$\eta_2 \in [0, 1]$. Furthermore, Mehamdia and Chaib [29] introduced a hybrid CG method that effectively integrates the strengths of various formulations with good convergence properties, where its conjugate parameter is defined as

$$\beta_k^{MDY^*} = \frac{\omega_k}{\rho_k} \frac{||H_{k+1}||^2 - \frac{\eta |H_{k+1}^T d_k| H_{k+1}^T H_k|}{||d_k|| ||H_k||}}{d_k^T (H_{k+1} - H_k) + \mu |H_{k+1}^T d_k|},$$

with $\omega_k = \frac{|H_{k+1}^T d_k|}{-H_k^T d_k}$, $\rho_k = 1 - \min\left\{0, \frac{v |H_{k+1}^T d_k| H_{k+1}^T H_k}{||H_{k+1}||^2 ||d_k|| ||H_k||}\right\}$, $v \geq 1$, $\mu > 1$, and $\eta \in [0, 1]$. The proposed variant significantly enhances numerical stability, accelerates convergence rates, and improves robustness compared to classical approaches. Recently, Arazm et al. [2] introduced another hybrid CG variant with the following CG parameter:

$$\beta_k^{\mathrm{DL+}} = \max\left\{\frac{H_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - t \frac{H_{k+1}^T s_k}{d_k^T y_k}.$$

This formulation has since been recognized as an effective strategy for balancing convergence guarantees and computational efficiency, particularly in settings where the objective function exhibits both convex and nonconvex behavior. For further recent advances in CG methods, the reader is referred to [25, 26, 28, 44].

In this context, the present work focuses on developing a novel CG algorithm specifically designed for training ANNs in the challenging domain of medical diagnosis. The proposed method aims to achieve superior convergence behavior, enhanced numerical stability, and improved classification accuracy compared to existing approaches. To further demonstrate its robustness and generalization capability, the algorithm is evaluated on two different medical classification problems, namely early breast cancer diagnosis and ECG signal classification for arrhythmia detection using the MIT-BIH dataset. The ECG task focuses on classifying heartbeat signals into multiple categories representing normal and abnormal cardiac rhythms.

4. The suggested conjugate gradient algorithm

4.1. The combined conjugate gradient formula and associated algorithm

Hybrid CG methods are characterized by their ability to combine the good theoretical properties of two or more methods, leading to improved descent property, global convergence guarantee, enhanced numerical stability, and accelerated convergence rates, especially when solving large-scale and nonconvex optimization problems. Many hybrid methods in the literature have demonstrated remarkable numerical performance compared to their classical counterparts. Based on this, our proposed hDL^{+} method employs the hybrid denominator $\max\{\|H_k\|^2, d_k^T y_k\}$ used in the JHJ method [43]. We also adopted the idea of multiplying β_k by the coefficient ρ_k from the MN method [21], which led to noticeable improvements in numerical results. Furthermore, we added the term $\mu|H_{k+1}^T d_k|$ to the denominator as used by Dai and Wen [13] to improve the descent property without any conditions on the line search used and to achieve numerical stability. In addition, we introduced a condition that allows the parameter β_k to take the value β_k^{DL+}, which is considered one of the most successful CG methods in terms of numerous results and convergence speed. Accordingly, the CG parameter used in our method is given as follows:

$$\beta_k^{hDL^{+}} = \begin{cases} \dfrac{1-\tau_k}{\rho_k}\dfrac{\|H_{k+1}\|^2 - \max\{\Phi_k, 0\}}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu|H_{k+1}^T d_k|} - t\dfrac{H_{k+1}^T s_k}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu|H_{k+1}^T d_k|}, & \text{if } H_{k+1}^T d_k \le 0 \text{ or } d_k^T y_k \le \|H_k\|^2, \\ \max\left\{\dfrac{H_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - t\dfrac{H_{k+1}^T s_k}{\max\left(\|H_k\|^2, d_k^T y_k\right)}, & \text{otherwise,} \end{cases} \tag{4.1}$$

where

$$\Phi_k = \frac{|H_{k+1}^T H_k| H_{k+1}^T H_k}{\|H_k\|^2}, \rho_k = 1 - \min\left\{0, \frac{\varsigma |H_{k+1}^T H_k| H_{k+1}^T H_k}{\|H_{k+1}\|^2 \|H_k\|^2}\right\}, \varsigma \ge 1, t \ge \frac{\varpi \|y_k\|^2}{s_k^T y_k}, \text{ and } \tau_k = \frac{|H_{k+1}^T d_k|}{-H_k^T d_k}.$$

For simplicity, the parameter $\beta_k^{hDL^+}$ can be expressed in the following compact form:

$$\beta_k^{hDL^+} = \begin{cases} \beta_k^* - t\dfrac{H_{k+1}^T s_k}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu|H_{k+1}^T d_k|}, & \text{if } H_{k+1}^T d_k \le 0 \text{ or } d_k^T y_k \le \|H_k\|^2, \\ \beta_k^{DL^+}, & \text{otherwise,} \end{cases}$$

and the search direction d_k is computed recursively by

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k^{\mathrm{hDL}^+} d_k. \tag{4.2}$$

Algorithm 1 below outlines the main steps of the hDL$^+$ method, which employs the strong Wolfe line search to determine a suitable step size α_k.

Algorithm 1: The hDL$^+$ algorithm

Step 0: (Initialization) Choose an initial point $x_0 \in \mathbb{R}^n$ and the parameters $0 < \delta < \sigma < 1$. Compute $f(x_0)$ and H_0. Set $d_0 = -H_0$.

Step 1: If $\|H_k\|_\infty \le 10^{-6}$, then stop. Otherwise, proceed to the next step.

Step 2: Determine the step size α_k satisfying the strong Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k H_k^T d_k, \tag{4.3}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \le -\sigma H_k^T d_k, \tag{4.4}$$

and compute the new point $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: β_k **computation:** Compute β_k using formula (4.1).

Step 4: **Search direction update:** Generate the search direction using formula (4.2).

Step 5: Set $k = k + 1$ and go back to Step 1.

4.2. The sufficient descent property

Before studying the convergence of the hDL$^+$ method, we first establish in the following lemma the descent property of the generated directions.

Theorem 4.1. *Consider the sequences $\{H_k\}_{k\ge0}$ and $\{d_k\}_{k\ge0}$ generated by the hDL$^+$ algorithm. For any index $k \ge 0$ with $H_k \ne 0$, the following descent condition holds:*

$$H_k^T d_k \le -\phi \|H_k\|^2, \ \textit{for all } k \ge 0, \tag{4.5}$$

where $\phi = 1 - \frac{1}{4\omega}$ and $\omega > \frac{1}{4}$.

Proof. We establish the result through mathematical induction, then we have

$$H_0^T d_0 = -\|H_0\|^2 \le -\left(1 - \frac{1}{4\omega}\right)\|H_0\|^2.$$

This satisfies both the required inequality and the descent condition. Assume the inequality (4.5) holds for some index k. Consequently, we have

$$H_k^T d_k < 0.$$

Using condition (4.4), it follows that

$$d_k^T (H_{k+1} - H_k) \geq (1 - \sigma)\left(-H_k^T d_k\right) > 0. \tag{4.6}$$

Before proceeding to prove that the inequality holds for $k+1$, we first establish an upper bound for the parameter β_k^*. Let ξ_k be the angle between H_k and H_{k+1}, so

$$\cos \xi_k = \frac{H_{k+1}^T H_k}{\|H_{k+1}\| \|H_k\|}.$$

From the condition (4.4), we know

$$0 < 1 - \sigma \leq 1 - \tau_k = 1 - \frac{\left|H_{k+1}^T d_k\right|}{-H_k^T d_k} < 1.$$

From the definition of β_k^*, it is clear that

$$\beta_k^* = \frac{[1 - \tau_k]\,(1 - \max(|\cos \xi_k| \cos \xi_k, 0))\,\|H_{k+1}\|^2}{[1 - \min(0, \varsigma |\cos \xi_k| \cos \xi_k)]\left[\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|\right]}.$$

Given $\varsigma \geq 1$, it can be shown that $0 \leq \frac{1-\tau_k}{\rho_k} < 1$. This is verified by two cases:

If $\cos \xi_k < 0$, we get

$$0 \leq \frac{1 - \tau_k}{\rho_k} = \frac{1 - \tau_k}{1 - \varsigma |\cos \xi_k| \cos \xi_k} \leq \frac{1}{1 - \varsigma |\cos \xi_k| \cos \xi_k} < 1, \tag{4.7}$$

then

$$0 \leq \beta_k^* = \frac{1 - \tau_k}{1 - \varsigma |\cos \xi_k| \cos \xi_k} \frac{\|H_{k+1}\|^2}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|} \leq \frac{\|H_{k+1}\|^2}{\max\left(\|H_k\|^2, d_k^T y_k\right)}. \tag{4.8}$$

If $\cos \xi_k \geq 0$, then

$$0 \leq \frac{1 - \tau_k}{\rho_k} = \frac{1 - \tau_k}{1 - \min(0, \varsigma |\cos \xi_k| \cos \xi_k)} = 1 - \tau_k < 1. \tag{4.9}$$

Hence,

$$0 \leq \beta_k^* \leq \frac{\|H_{k+1}\|^2}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|} \leq \frac{\|H_{k+1}\|^2}{\max\left(\|H_k\|^2, d_k^T y_k\right)}. \tag{4.10}$$

Combining (4.8) and (4.10), the following inequality holds regardless of the sign of $\cos \xi_k$:

$$0 \leq \beta_k^* \leq \frac{\|H_{k+1}\|^2}{\max\left(\|H_k\|^2, d_k^T y_k\right)}. \tag{4.11}$$

Having established this bound, we now prove that the descent inequality holds at iteration k+1 by considering the following three cases:

Case (i)**:** $H_{k+1}^T d_k \leq 0$. Using the positivity of β_k^* from (4.11), we obtain

$$H_{k+1}^T d_{k+1} = -\|H_{k+1}\|^2 + \beta_k^* H_{k+1}^T d_k - t\alpha_k \frac{\left(H_{k+1}^T d_k\right)^2}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|}$$

$$\leq \quad -\|H_{k+1}\|^2 + \beta_k^* H_{k+1}^T d_k \leq -\|H_{k+1}\|^2 \leq -\phi \|H_{k+1}\|^2 .$$

Case (*ii*)**:** $H_{k+1}^T d_k > 0$ and $y_k^T d_k \leq \|H_k\|^2$. In this situation, we have

$$0 < H_{k+1}^T d_k \leq H_k^T d_k + \|H_k\|^2 , \tag{4.12}$$

and from Eq (4.11) we have

$$\begin{aligned} H_{k+1}^T d_{k+1} &= -\|H_{k+1}\|^2 + \beta_k^* H_{k+1}^T d_k - t\alpha_k \frac{\left(H_{k+1}^T d_k\right)^2}{\max\left(\|H_k\|^2 , d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|} \\ &\leq -\|H_{k+1}\|^2 + \beta_k^* H_{k+1}^T d_k \\ &\leq -\|H_{k+1}\|^2 + \frac{\|H_{k+1}\|^2 H_{k+1}^T d_k}{\|H_k\|^2}. \end{aligned}$$

Applying (4.12) leads to

$$\begin{aligned} H_{k+1}^T d_{k+1} &\leq -\|H_{k+1}\|^2 + \frac{\|H_{k+1}\|^2 \left[H_k^T d_k + \|H_k\|^2\right]}{\|H_k\|^2} \\ &= \frac{\|H_{k+1}\|^2 H_k^T d_k}{\|H_k\|^2} \leq \left(-\phi \|H_k\|^2\right) \frac{\|H_{k+1}\|^2}{\|H_k\|^2} \\ &= -\phi \|H_{k+1}\|^2 . \end{aligned}$$

Case (*iii*)**:** $H_{k+1}^T d_k > 0$ and $d_k^T y_k > \|H_k\|^2$. From (4.1) and (4.2), we can express

$$H_{k+1}^T d_{k+1} = -\|H_{k+1}\|^2 + \left[\max\left\{\frac{H_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - \eta \frac{\|y_k\|^2}{s_k^T y_k} \frac{H_{k+1}^T s_k}{d_k^T y_k}\right] H_{k+1}^T d_k.$$

Now, if $H_{k+1}^T y_k \leq 0$, then we get

$$\begin{aligned} H_{k+1}^T d_{k+1} &= -\|H_{k+1}\|^2 - \eta \frac{\|y_k\|^2 \left(H_{k+1}^T s_k\right)^2}{\left(s_k^T y_k\right)^2} \\ &\leq -\|H_{k+1}\|^2 \\ &\leq -\phi \|H_{k+1}\|^2 . \end{aligned}$$

When $H_{k+1}^T y_k > 0$, the result follows by applying Theorem 1 of ([3]). □

5. Global convergence

The global convergence feature is essential for any CG method. To establish it, we assume that:

Assumption 1. *The level set* $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ *is bounded.*

Assumption 2. *In some open convex neighbourhood* $\mathcal{N}$ *of* S*, the function* f *is continuously differentiable, and its gradient is Lipschitz continuous, namely, there exists a constant* $L > 0$ *such that*

$$\|H(x) - H(x')\| \leq L \|x - x'\| \text{ for all } x, x' \in \mathcal{N}. \tag{5.1}$$

From assumptions 1 and 2, we deduce that for all $x \in S$, there exists $\vartheta, \Gamma > 0$, for which

$$\|x\| \leq \vartheta, \tag{5.2}$$

and

$$\| H(x) \| \leq \Gamma. \tag{5.3}$$

To establish the global convergence of the hDL$^+$ method, we rely on the well-known Zoutendijk condition [46].

Lemma 5.1. *Consider any iterative method of the form* (3.1)*, where d_k is a descent direction and α_k satisfies the Wolfe conditions* (4.3) *and* (4.4)*. Then, under the above assumptions, the so-called Zoutendijk condition holds:*

$$\sum_{k=0}^{\infty} \frac{(H_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{5.4}$$

In our analysis, the following property, namely Property (*), plays an important role.
Property (*). Suppose that the above assumptions hold and assume that

$$0 < \gamma \leq \|H_k\| \leq \Gamma \text{ for all } k \geq 0. \tag{5.5}$$

A CG method in the form (3.1), (3.2) is said to satisfy Property (*) if there exist constants $b > 1$ and $\lambda > 0$ such that, for all k,

$$|\beta_k| \leq b \text{ and } \|x_{k+1} - x_k\| \leq \lambda,$$

which implies that

$$|\beta_k| \leq \frac{1}{2b}.$$

The following lemma proves that the proposed hDL$^+$ method satisfies Property (*).

Lemma 5.2. *Under the above considerations, the new β_k^{hDL+} satisfies Property* (*).

Proof. We aim to show that the hDL$^+$ method satisfies Property (*). To do this, we assume that $\|x_{k+1} - x_k\| \leq \lambda$ and examine β_k^{hDL+} in two cases depending on the sign of $H_{k+1}^T H_k$.
Case(i): If $\beta_k^{hDL^+} = \beta_k^*$, and $H_{k+1}^T H_k > 0$. Using (4.7), (4.9), and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \left|\beta_k^{hDL^+}\right| &\leq \frac{\left|\|H_{k+1}\|^2 - \frac{|H_{k+1}^T H_k|}{\|H_k\|^2} H_{k+1}^T H_k\right|}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|} + t \frac{\left|H_{k+1}^T s_k\right|}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu \left|H_{k+1}^T d_k\right|} \\ &\leq \frac{\left|\|H_{k+1}\|^2 - \frac{|H_{k+1}^T H_k|}{\|H_k\|^2} H_{k+1}^T H_k\right|}{d_k^T y_k} + t \frac{\left|H_{k+1}^T s_k\right|}{d_k^T y_k}. \end{aligned}$$

By the Cauchy-Schwarz inequality, we obtain

$$\left| H_{k+1}^T \left(H_{k+1} - \frac{\left|H_{k+1}^T H_k\right|}{\|H_k\|^2} H_k \right) \right| \leq \|H_{k+1}\| \left(\|H_{k+1} - H_k\| + \left\| H_k - \frac{\left|H_{k+1}^T H_k\right|}{\|H_k\|^2} H_k \right\| \right)$$

$$
\begin{aligned}
&\leq \left| \|H_{k+1}\| \left(\|H_{k+1} - H_k\| + \|H_k\| \left\| \frac{H_k^T H_k - H_{k+1}^T H_k}{\|H_k\|^2} \right\| \right) \right| \\
&\leq \left| \|H_{k+1}\| \left(\|H_{k+1} - H_k\| + \|H_k\| \left\| \frac{(H_k - H_{k+1})^T H_k}{\|H_k\|^2} \right\| \right) \right| \\
&= \left| \|H_{k+1}\| \left(\|H_{k+1} - H_k\| + \|H_k\|^2 \left\| \frac{H_k - H_{k+1}}{\|H_k\|^2} \right\| \right) \right| \\
&= 2 \|H_{k+1}\| \|H_{k+1} - H_k\| .
\end{aligned}
$$

In this situation, from (4.5), (4.6), and (5.5), we have

$$
d_k^T y_k \geq (1 - \sigma) \left(-H_k^T d_k \right) \geq (1 - \sigma) \phi \Gamma^2 > 0. \tag{5.6}
$$

Using the result of bounding the numerator (5.3), and (5.6),

$$
\begin{aligned}
\left| \beta_k^{hDL^+} \right| &\leq \frac{2L \|H_{k+1}\| \|s_k\|}{d_k^T y_k} + T \frac{\|H_{k+1}\| \|s_k\|}{d_k^T y_k} \\
&\leq \lambda \frac{\Gamma (2L + T)}{(1 - \sigma) \phi \gamma^2} = \lambda b_1, \text{ in which } b_1 = \frac{\Gamma (2L + T)}{(1 - \sigma) \phi \gamma^2}.
\end{aligned}
$$

If $\beta_k^{hDL^+} = \beta_k^*$ and $H_{k+1}^T H_k \leq 0$. For this case, since $H_{k+1}^T y_k = \|H_{k+1}\|^2 - H_{k+1}^T H_k \geq \|H_{k+1}\|^2$, we have

$$
\begin{aligned}
\left| \beta_k^{hDL^+} \right| &\leq \left| \frac{\|H_{k+1}\|^2}{d_k^T y_k} - t \frac{H_{k+1}^T s_k}{d_k^T y_k} \right| \\
&\leq \frac{\left| H_{k+1}^T y_k \right|}{d_k^T y_k} + t \frac{\left| H_{k+1}^T s_k \right|}{d_k^T y_k} \\
&\leq \frac{\|H_{k+1}\| \|H_{k+1} - H_k\|}{d_k^T y_k} + t \frac{\|H_{k+1}\| \|s_{k+1}\|}{d_k^T y_k}.
\end{aligned}
$$

From $\|s_k\| \leq \lambda$ and (5.3), we get

$$
\begin{aligned}
\left| \beta_k^{hDL^+} \right| &\leq \frac{L \|H_{k+1}\| \|s_k\|}{d_k^T y_k} + T \frac{\|H_{k+1}\| \|s_k\|}{d_k^T y_k} \\
&\leq \lambda \frac{\Gamma (L + T)}{(1 - \sigma) \phi \gamma^2} = \lambda b_2.
\end{aligned}
$$

Case (ii): $\beta_k^{hDL^+} = \beta_k^{DL+}$. Then,

$$
\begin{aligned}
\left| \beta_k^{hDL^+} \right| &\leq \left| \frac{\left| H_{k+1}^T y_k \right|}{d_k^T (H_{k+1} - H_k)} - t \frac{\left| H_{k+1}^T s_k \right|}{d_k^T (H_{k+1} - H_k)} \right| \\
&\leq \frac{\left| H_{k+1}^T y_k \right|}{d_k^T (H_{k+1} - H_k)} + T \frac{\left| H_{k+1}^T s_k \right|}{d_k^T (H_{k+1} - H_k)} \\
&\leq \frac{\|H_{k+1}\| \|H_{k+1} - H_k\|}{(1 - \sigma) \phi \Gamma^2} + T \frac{\|H_{k+1}\| \|s_k\|}{(1 - \sigma) \phi \Gamma^2}.
\end{aligned}
$$

Also, by using $\|s_k\| \leq \lambda$ and (5.3),

$$\begin{aligned} \left|\beta_k^{hDL^+}\right| &\leq \frac{L\|H_{k+1}\|\|s_k\|}{(1-\sigma)\phi\Gamma^2} + T\frac{\|H_{k+1}\|\|s_k\|}{(1-\sigma)\phi\Gamma^2} \\ &\leq \lambda\frac{\Gamma(L+T)}{(1-\sigma)\phi\gamma^2} = \lambda b_2, \end{aligned}$$

in which $b_2 = \frac{\lambda(\Gamma+T)}{(1-\sigma)\phi\gamma^2}$ can be appropriately set to get $b_2 > 1$.

From both cases, we can choose $b = \Gamma\max(b_1, b_2)$ and set $\lambda = \frac{1}{2b\Gamma\max(b_1,b_2)}$. Then, whenever $\|x_{k+1} - x_k\| \leq \lambda$, we have $\left|\beta_k^{hDL^+}\right| \leq \frac{1}{2b}$. Thus, the hDL$^+$ method satisfies Property (*). □

We now state the result of Dai et al. [11], which is also crucial for proving global convergence.

Lemma 5.3. *Consider any iterative method of the form* (3.1)–(3.2) *with descent directions d_k, where the step size α_k meets the strong Wolfe conditions* (4.3)–(4.4)*. Under assumptions 1 and 2, if*

$$\sum_{k\geq 0}\frac{1}{\|d_k\|^2} = \infty,$$

then

$$\liminf_{k\to\infty}\|H_k\| = 0.$$

We next prove the following lemma:

Lemma 5.4. *Let $\{H_k\}_{k\in\mathbb{N}}$ and $\{d_k\}_{k\in\mathbb{N}}$ be the sequences produced by the hDL$^+$ algorithm and assume that Assumption 1 holds. Suppose that there exists a constant $\gamma > 0$ such that*

$$\|H_k\| \geq \gamma \text{ for all } k \geq 0,$$

then $d_k \neq 0$ and

$$\sum_{k\geq 0}\|\eta_{k+1} - \eta_k\|^2 < \infty,$$

where $\eta_k = \frac{d_k}{\|d_k\|}$.

Proof. First, we observe that $d_k \neq 0$ for all k. Indeed, if $d_k = 0$ for some k, then from (3.2), we would have $H_k = 0$, which contradicts $\|H_k\| \geq \gamma > 0$. Hence, $\eta_k = \frac{d_k}{\|d_k\|}$ is well defined. From the Zoutendijk condition (5.4) and the relation (4.5), we have

$$\sum_{k=0}^{\infty}\frac{(H_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Using (4.5), we know that $H_k^T d_k \leq -\phi\|H_0\|^2 \leq -\phi\gamma^2$. Therefore, $\left(H_k^T d_k\right)^2 \geq \phi^2\gamma^4$. Substituting into the Zoutendijk condition gives

$$\sum_{k=0}^{\infty}\frac{\phi^2\gamma^4}{\|d_k\|^2} \leq \sum_{k=0}^{\infty}\frac{(H_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Consequently,

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} < \infty. \tag{5.7}$$

We now split the expression (4.1) for $\beta_k^{hDL^+}$ into two parts as follows:

$$\beta_k^{hDL^+} = \beta_k^{(1)} + \beta_k^{(2)},$$

where

$$\beta_k^{(1)} = \begin{cases} \beta_k^*, \text{ if } H_{k+1}^T d_k \leq 0 \text{ or } d_k^T y_k \leq \|H_k\|^2, \\ \max\left\{\frac{H_{k+1}^T y_k}{d_k^T y_k}, 0\right\}, \text{otherwise,} \end{cases}$$

and

$$\beta_k^{(2)} = \begin{cases} -t\frac{H_{k+1}^T s_k}{\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu\left|H_{k+1}^T d_k\right|}, \text{ if } H_{k+1}^T d_k \leq 0 \text{ or } d_k^T y_k \leq \|H_k\|^2, \\ -t\frac{H_{k+1}^T s_k}{\max\left(\|H_k\|^2, d_k^T y_k\right)}, \text{otherwise.} \end{cases}$$

Define $r_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|}$ and $\delta_k = \frac{\beta_k^{(1)}\|d_k\|}{\|d_{k+1}\|}$, where $v_{k+1} = -H_{k+1} + \beta_k^{(2)} d_k$. From the search direction formula (3.2), we have

$$d_{k+1} = \upsilon_{k+1} + \beta_k^{(1)} d_k.$$

Dividing both sides by $\|d_{k+1}\|$ yields

$$\eta_{k+1} = \frac{\upsilon_{k+1}}{\|d_{k+1}\|} + \frac{\beta_k^{(1)}\|d_k\|\eta_k}{\|d_{k+1}\|} = r_{k+1} + \delta_k \eta_k. \tag{5.8}$$

From (5.8) and the fact that $\|\eta_{k+1}\| = \|\eta_k\| = 1$, we obtain

$$\|r_{k+1}\| = \|\eta_{k+1} - \delta_k \eta_k\| = \|\delta_k \eta_{k+1} - \eta_k\|.$$

Applying the triangle inequality, we get

$$\|\eta_{k+1} - \eta_k\| \leq \|\eta_{k+1} - \delta_k \eta_{k+1}\| + \|\delta_k \eta_{k+1} - \eta_k\| = 2\|r_{k+1}\|.$$

We now proceed to bound $\|v_{k+1}\|$. From its definition, we have

$$\|v_{k+1}\| \leq \|H_{k+1}\| + \left|\beta_k^{(2)}\right| \|d_k\|.$$

From (5.6), we have

$$\frac{\left|H_k^T d_k\right|}{d_k^T y_k} \leq \frac{\sigma}{(1-\sigma)},$$

and

$$\max\left(\|H_k\|^2, d_k^T y_k\right) + \mu\left|H_{k+1}^T d_k\right| \geq d_k^T y_k,$$

then we can further bound $\left|\beta_k^{(2)}\right|$ as

$$\left|\beta_k^{(2)}\right| \leq t\frac{\left|H_{k+1}^T s_k\right|}{d_k^T y_k} = t\frac{\alpha_k\left|H_{k+1}^T d_k\right|}{d_k^T y_k}.$$

Given that $\|s_k\| \leq \lambda$ and $\|H_k\| \leq \Gamma$, we obtain

$$\|v_{k+1}\| \leq \Gamma + t\frac{\sigma\alpha_k \|d_k\|}{(1-\sigma)} \leq \Gamma + \frac{\sigma \|s_k\| T}{(1-\sigma)} \leq \Gamma + \frac{\sigma\lambda T}{(1-\sigma)}.$$

From the definition of r_{k+1},

$$\|r_{k+1}\|^2 = \frac{\|v_{k+1}\|^2}{\|d_{k+1}\|^2} \leq \frac{C^2}{\|d_{k+1}\|^2}, \text{ where } c = \Gamma + \frac{\sigma\lambda T}{(1-\sigma)}.$$

From (5.7), we have

$$\sum_{k=0}^{\infty} \|\eta_{k+1} - \eta_k\|^2 \leq 4 \sum_{k=0}^{\infty} \|r_{k+1}\|^2 \leq 4C^2 \sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}\|^2} < \infty.$$

This completes the proof of Lemma 5.4.

□

The following theorem, due to Babaie-Kafaki [3], is crucial for proving the global convergence.

Lemma 5.5. *Consider an iterative method defined by* (3.1) *and* (3.2)*, equipped with any line search that ensures the following:* $\{x_k\} \subset S$*, the Zoutendijk condition* (5.4)*, and the sufficient descent condition* (4.5)*. Further, assume that the method satisfies Property (*) and that condition* (5.5) *holds. Then, under assumptions 1 and 2, there exists a constant* $\tau > 0$ *such that, for any* $\triangle \in \mathbb{N}^*$ *and any index* k_0*, there exists a* $k \geq k_0$*, for which*

$$size\ \mathfrak{R}_{k,\triangle}^{\tau} > \frac{\triangle}{2},$$

where

$$\mathfrak{R}_{k,\triangle}^{\tau} = \{i \in \mathbb{N} : k \leq i \leq k + \triangle - 1, \|x_{i+1} - x_i\| > \tau\}$$

and size $\mathfrak{R}_{k,\triangle}^{\tau}$ *denotes the cardinality of the set* $\mathfrak{R}_{k,\triangle}^{\tau}$.

We now establish our main results, which concern the global convergence of the hDL$^+$ algorithm.

Theorem 5.1. *Let the sequences* $\{H_k\}_{k\geq 0}$, $\{d_k\}_{k\geq 0}$ *be generated by the hDL*$^+$ *algorithm. Under the above considerations, the hDL*$^+$ *algorithm converges in the sense that*

$$\liminf_{k\to\infty} \|H_k\| = 0.$$

Proof. Assume the contrary, namely that (5.5) holds. Consequently, the conditions of Lemmas 4 and 5 are satisfied. Defining $\eta_k = \frac{d_k}{\|d_k\|}$. As before, we have, for any two indices l, k, with $l \geq k$,

$$\begin{aligned} x_{l+1} - x_k &= \sum_{i=k}^{l} \|x_{i+1} - x_i\| \eta_i \\ &= \sum_{i=k}^{l} \|s_i\| \eta_k + \sum_{i=k}^{l} \|s_i\| (\eta_i - \eta_k). \end{aligned}$$

Now, applying the norm yields

$$\sum_{i=k}^{l} \|s_i\| \leq \|x_{l+1} - x_k\| + \sum_{i=k}^{l} \|s_i\| \|\eta_i - \eta_k\|.$$

From inequality (5.2), it follows that

$$\sum_{i=k}^{l} \|s_i\| \leq 2\vartheta + \sum_{i=k}^{l} \|s_i\| \|\eta_i - \eta_k\|. \tag{5.9}$$

Let $\tau > 0$ be as in Lemma 5.5. Using the same notation, put $\triangle = [\frac{8\vartheta}{\tau}]$, where $\frac{8\vartheta}{\tau} \leq \triangle < \frac{8\vartheta}{\tau} + 1$ and $\triangle \in \mathbb{N}^*$. By Lemma 5.4, we can choose k_0 such that

$$\sum_{i \geq k_0} \|\eta_{i+1} - \eta_i\|^2 \leq \frac{1}{4\triangle}. \tag{5.10}$$

With this $\triangle$ and k_0 fixed, Lemma 5.5 allows us to find an index $k > k_0$ such that

$$size\ \mathfrak{R}^{\tau}_{k,\triangle} > \frac{\triangle}{2}. \tag{5.11}$$

Next, for any index $i \in [k, k + \triangle - 1]$, we have, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\eta_i - \eta_k\| &\leq \sum_{j=k}^{i-1} \left\|\eta_{j+1} - \eta_j\right\| \\ &\leq (i-k)^{\frac{1}{2}} \left(\sum_{j=k}^{i-1} \left\|\eta_{j+1} - \eta_j\right\|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

From (5.10), we get

$$\|\eta_i - \eta_k\| \leq \frac{1}{2}. \tag{5.12}$$

By (5.9), (5.11), and (5.12), with $l = k + \triangle - 1$, we have

$$2\vartheta \geq \frac{1}{2} \sum_{i=k}^{k+\triangle-1} \|s_i\| > \frac{\tau}{2} size\ \mathfrak{R}^{\tau}_{k,\triangle} > \frac{\triangle\tau}{4}. \tag{5.13}$$

From (5.13), we have $\triangle < \frac{8\vartheta}{\tau}$, which is a contradiction with the definition of $\triangle$. Therefore,

$$\liminf_{k\to\infty} \|H_k\| = 0.$$

Thus, we complete the proof. □

6. Computational experiments

All the numerical experiments are conducted in MATLAB 2023 and executed on a PC with a 2.5 GHz processor and 16 GB of RAM, running the Windows 11 operating system. In all tests, the hDL$^+$ method is implemented using the following parameter settings: $\mu = 1$, $\varsigma = 2$, and $\varpi = 0.5$. For the strong Wolfe line conditions, the parameters are set to $\delta = 10^{-2}$, $\sigma = 10^{-1}$.

6.1. Comprehensive performance analysis for early breast cancer prediction

Table 1 reports the quantitative comparison between the classical HS-CG method and the proposed hDL$^+$ CG algorithm for neural network training in early breast cancer prediction.

Table 1. Performance comparison of ANN training algorithms for breast cancer prediction.

Method	Accuracy (%)	Precision	Recall	F1-score	MSE
HS-CG	95.78	0.97	0.91	0.94	0.0211
hDL$^+$	98.24	0.99	0.97	0.98	0.00997

6.1.1. Overall classification accuracy

With an accuracy of 98.24%, the suggested hDL$^+$ algorithm outperforms the HS-CG method, which reaches 95.78%. This indicates a 2.46% absolute improvement. Even slight advances in medical diagnosis are clinically relevant since each percentage point equates to accurately identifying more patients, even though this numerical difference may seem minor.

The increased accuracy suggests that the optimization process is guided toward a more representative solution of the underlying nonlinear classification boundary by the search directions produced by hDL$^+$.

Precision analysis: The accuracy of cancerous predictions is measured by precision. The precision of the hDL$^+$ approach is 0.99, while that of the HS-CG method is 0.97. A decrease in false positive diagnoses is implied by this improvement. Reducing false positives in practical screening systems helps cut down on needless biopsies, patient anxiety, and extra clinical expenses.

Recall (sensitivity) and clinical significance: In cancer prediction tasks, recall is perhaps the most important parameter. The HS-CG approach achieves a recall of 0.91, while the suggested strategy obtains 0.97. This translates to a relative improvement in the detection of malignant patients of about 6.6%. From a clinical standpoint, increasing recollection immediately lowers false negatives, which is crucial as overlooked malignant cases can cause treatment delays and have a substantial impact on survival rates. As a result, the hDL$^+$ algorithm exhibits better diagnostic safety.

F1-Score and class balance: Precision and recall are balanced by the F1-score, which rises from 0.94 (HS-CG) to 0.98 (hDL$^+$). This suggests that the suggested approach strikes a more steady balance between identifying cancerous tumours and preventing benign sample misclassification. Additionally, the improvement implies improved resilience to the dataset's slight class imbalance.

Training performance and convergence behavior: In comparison to the 0.0211 acquired by the HS-CG approach, the final performance (MSE) value attained by hDL$^+$ is 0.00997. This translates into a final goal function value decrease of more than 52%.

This decline suggests that the suggested CG formulation enhances curvature exploitation during optimization and generates more effective descent directions. Compared to the HS-CG method, hDL$^+$ exhibits more stable gradient behavior, faster training error reduction, and smoother convergence curves, as shown in Figure 2.

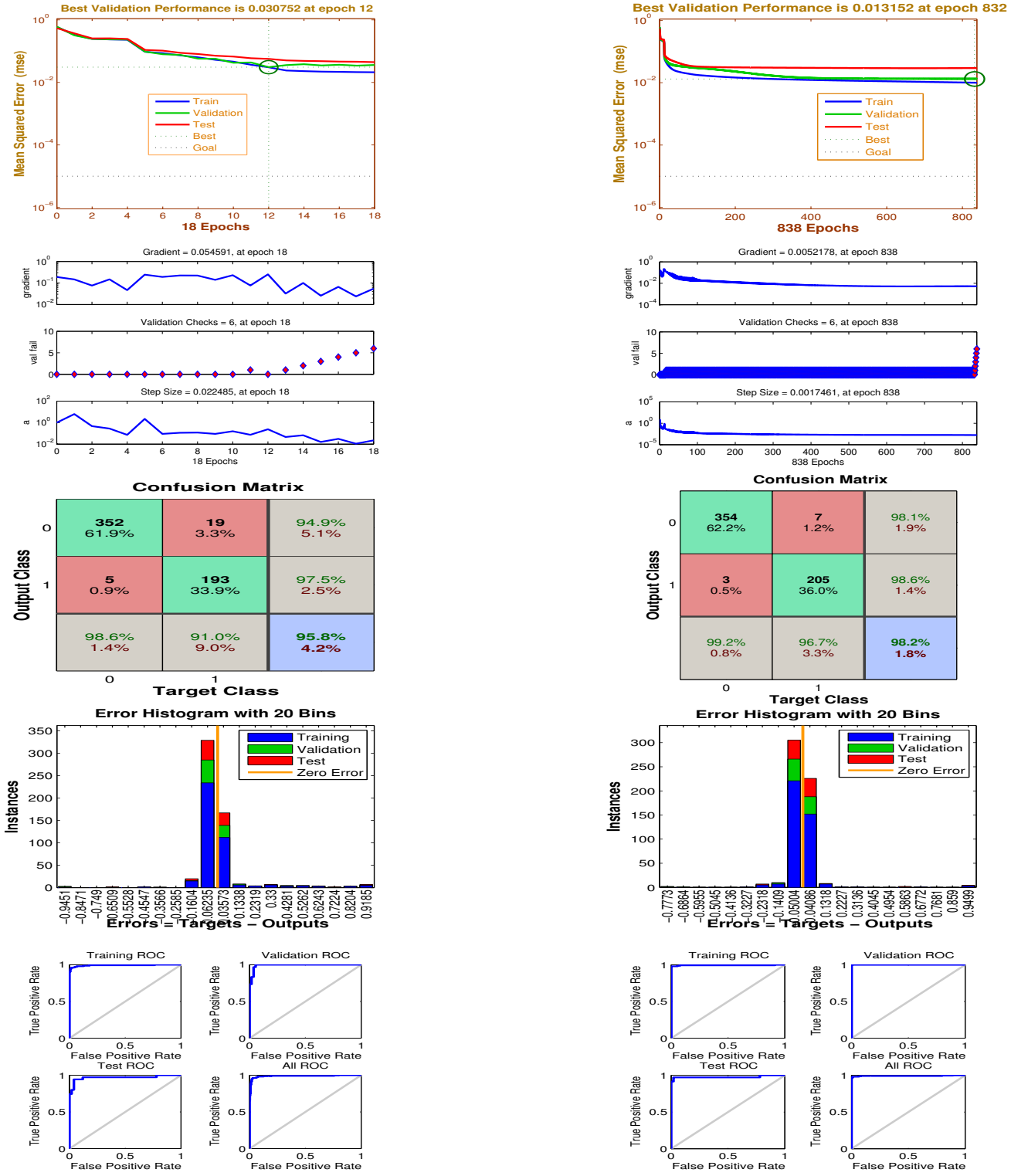

Figure 2. Comparison of training performance, convergence behavior, confusion matrices, and ROC curve analysis: HS-CG vs. proposed hDL$^+$ method.

Better numerical stability and better management of the nonlinear error surface related to neural network training are suggested by the improved convergence behavior.

Confusion matrix interpretation: The superiority of the suggested strategy is further shown by the confusion matrices in Figure 2. Because the hDL$^+$ method produces fewer false positives and false negatives, the distribution of malignant and benign classes is more evenly distributed.

Collectively, the experimental results demonstrate that the hDL$^+$ CG algorithm significantly enhances:

- Predictive accuracy,
- Sensitivity to malignant cases,
- Training convergence efficiency,
- Numerical stability of the optimization process.

These enhancements demonstrate that the suggested approach is both clinically significant and computationally effective, making it ideal for extensive medical decision-support systems.

6.2. Comprehensive performance analysis for ECG classification

Table 2 reports the quantitative comparison between the classical HS-CG method and the proposed hDL$^+$ CG algorithm for neural network training in ECG classification.

Table 2. Comprehensive performance comparison between HS-CG and hDL$^+$ on ECG dataset including classification metrics, AUC-ROC, error, and training time.

Metric	HS-CG	hDL$^+$
Overall performance		
Accuracy (%)	78.31	80.45
Precision	0.78	0.80
Recall	0.80	0.81
F1-score	0.78	0.80
MSE (Performance)	0.0665	0.0605
Training Time	0:02:51	0:01:13
Class-wise results		
Class 0 (P/R/F1)	0.73 / 0.58 / 0.65	0.76 / 0.63 / 0.69
Class 1 (P/R/F1)	0.83 / 0.86 / 0.84	0.85 / 0.90 / 0.87
Class 2 (P/R/F1)	0.56 / 0.84 / 0.67	0.62 / 0.84 / 0.71
Class 3 (P/R/F1)	0.89 / 0.86 / 0.87	0.89 / 0.86 / 0.88
Class 4 (P/R/F1)	0.91 / 0.85 / 0.88	0.91 / 0.84 / 0.87
AUC-ROC results		
Mean AUC	0.9024	0.9130
Class 0 AUC	0.8681	0.8879
Class 1 AUC	0.9199	0.9227
Class 2 AUC	0.7694	0.7964
Class 3 AUC	0.9759	0.9780
Class 4 AUC	0.9789	0.9801

6.2.1. Overall classification accuracy

Table 2 shows that the proposed hDL$^+$ method achieves better overall performance compared to the HS-CG method. The accuracy increases from 78.31% to 80.45%, representing a noticeable improvement in classification capability. Although the numerical gain appears moderate, even small improvements are important in medical signal classification tasks, where each correctly classified sample contributes to improved diagnostic reliability.

The improvement in accuracy indicates that the proposed hDL$^+$ update strategy provides more effective search directions during optimization, leading to a better approximation of the nonlinear decision boundaries in ECG signal classification.

Precision analysis: Precision reflects the reliability of positive predictions. The hDL$^+$ method achieves a precision of 0.80, compared to 0.78 for HS-CG. This improvement indicates a reduction in false positive predictions, which is important in ECG-based diagnosis systems to avoid incorrect abnormality detection.

Recall (sensitivity) analysis: Recall increases slightly from 0.80 (HS-CG) to 0.81 (hDL$^+$). This improvement demonstrates that the proposed method is more effective in identifying true positive ECG classes, which is critical for medical decision support systems where missed detections may lead to incorrect clinical interpretation.

F1-score and classification balance: The F1-score improves from 0.78 to 0.80, indicating a better balance between precision and recall. This suggests that the proposed method provides more stable classification performance across different ECG classes, particularly in the presence of class imbalance.

Area Under the receiver Operating-Characteristic Curve (AUC-ROC) analysis: The mean AUC increases from 0.9024 (HS-CG) to 0.9130 (hDL$^+$), confirming improved class separability. The per-class AUC values also show consistent improvements across all ECG classes, demonstrating the robustness of the proposed method in distinguishing between different cardiac conditions.

Training performance and convergence behavior: The MSE achieved by hDL$^+$ is 0.0605, compared to 0.0665 for HS-CG. This reduction indicates improved optimization efficiency and better convergence behavior.

In addition, the training time is significantly reduced from 0:02:51 to 0:01:13, which demonstrates that the proposed method not only improves accuracy but also enhances computational efficiency.

Confusion matrix interpretation: The confusion matrices in Figure 3 further confirm the superiority of the proposed method. The hDL$^+$ algorithm produces fewer misclassifications across all ECG classes, resulting in a more balanced and reliable classification performance.

Summary of results: Overall, the experimental results demonstrate that the hDL$^+$ CG algorithm improves:

- Classification accuracy,
- Class separability (AUC),
- Optimization convergence efficiency,
- Computational time,
- Stability of training performance.

These improvements confirm that the proposed method is both computationally efficient and effective for ECG-based medical decision support systems.

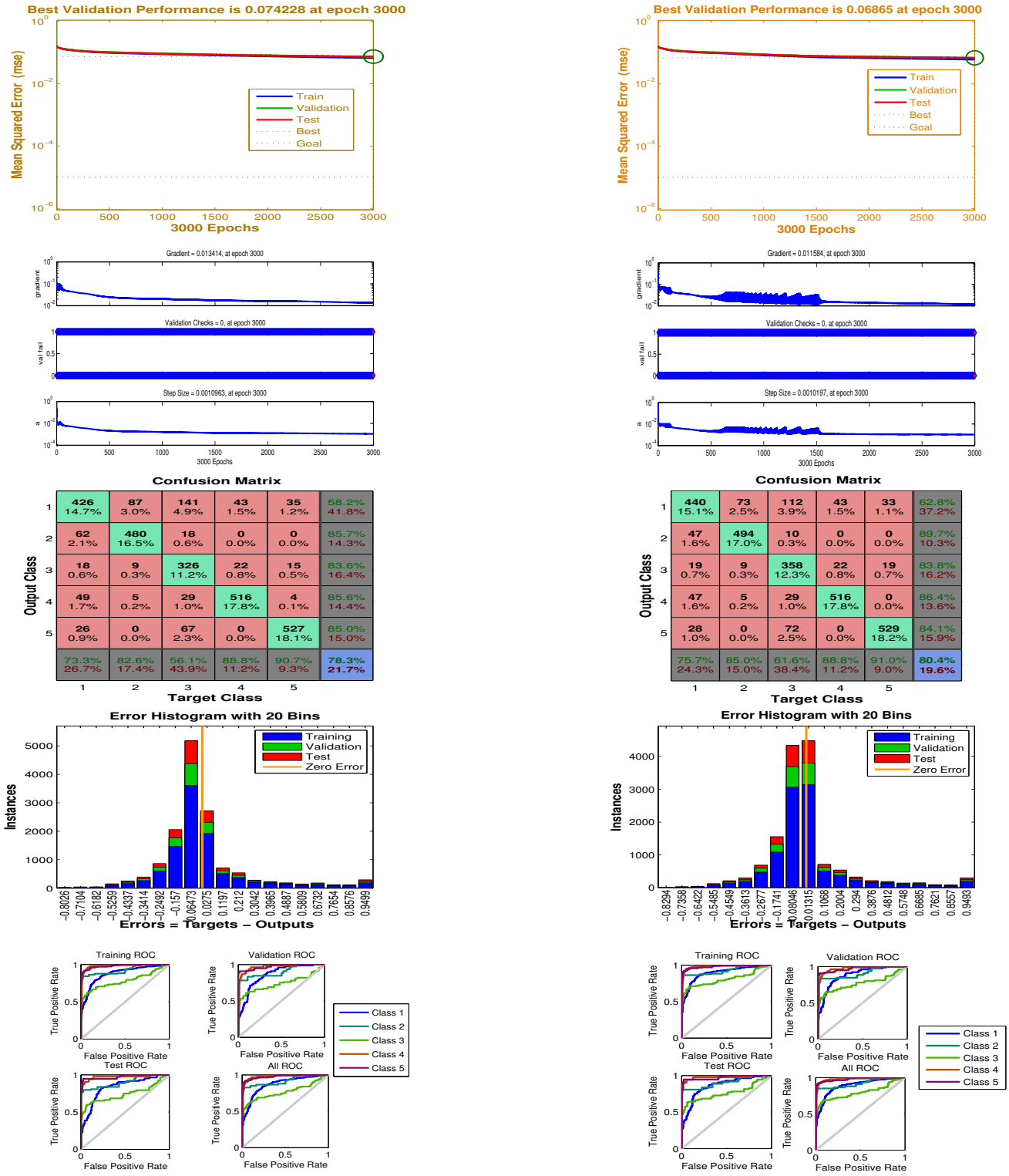

Figure 3. Comparison of training performance, convergence behavior, confusion matrices, and ROC curve analysis: HS-CG vs. proposed hDL^{+} method on ECG dataset.

6.3. Application on typical test functions

To validate the efficiency and robustness of the proposed method, we evaluated it on a set of test problems of varying dimensions from the CUTEr collection [1, 6]. To assess its performance, we compared it against four established CG methods: HS [20], DL$^+$ [2], MDY* (Modified Dai-Yuan) [29], and IFR (Improved Fletcher-Reeves) [30]. In our experiments, we utilized $t = \frac{\|y_k\|^2}{s_k^T y_k}$, which is the optimal choice suggested by Dai and Kou [12]. All methods were implemented under strong Wolfe line conditions with the parameter settings $\delta = 10^{-3}$, $\sigma = 10^{-1}$. An execution is deemed successful if the algorithm reaches the optimality condition $\|g_k\|_\infty \leq 10^{-6}$ within 2×10^3 iterations and a CPU time of 500 seconds; otherwise, the run is recorded as a failure.

Figures 4 and 5 illustrate the performance profiles of the compared methods with respect to the number of iterations and CPU time. This comparative analysis is conducted using the performance profiles of Dolan and Moré [15]. The top curve in the graphs corresponds to the method that achieves the highest success rate across the test problems within the factor t, as described in [15]. Let Q denote the set of solvers and P the set of test problems, with $n_q = |Q|$ and $n_p = |P|$ representing their respective cardinalities. For each problem $p \in P$ and solver $q \in Q$, let $\tau_{p,q}$ denote the performance metric (e.g., iteration count or CPU time) required by solver q to solve problem p. The performance ratio is defined as

$$r_{p,q} = \frac{\tau_{p,q}}{\min\{\tau_{p,i} : i \in Q\}}.$$

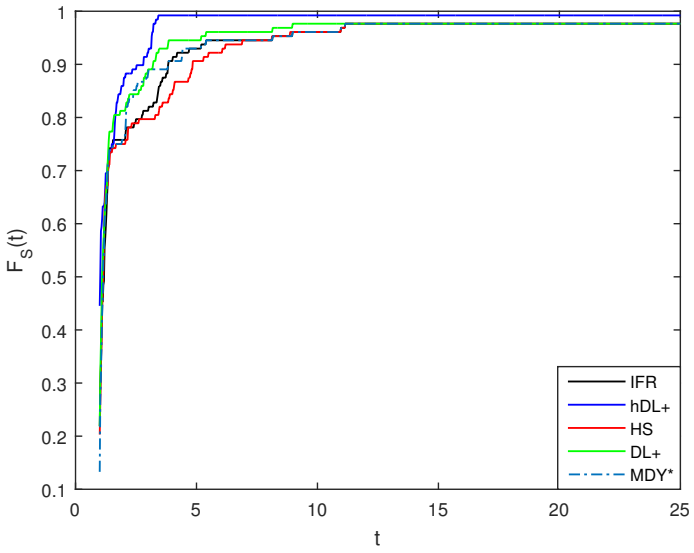

Figure 4. Performance profiles plot based on CPU time.

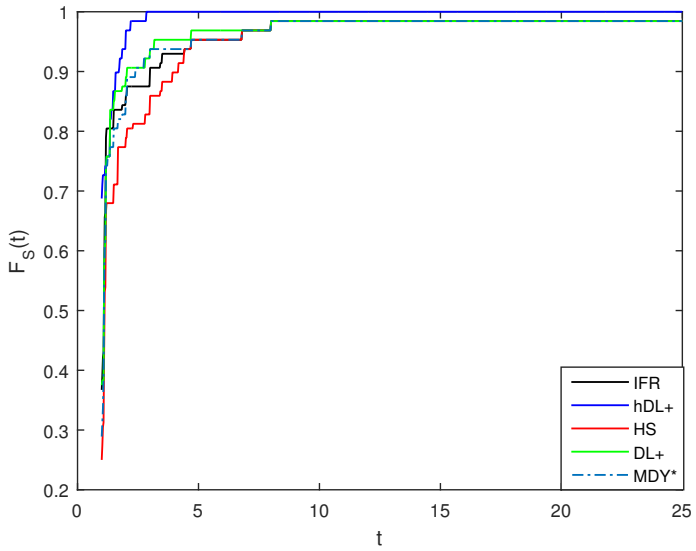

Figure 5. Performance profiles plot based on the number of iterations.

Assume there exists a parameter r_M such that $r_{p,q} \leq r_M$ for all $p \in P$ and $q \in Q$, where $r_{p,q} = r_M$ if and only if solver q fails to solve problem p. The global assessment of solver performance is provided by the performance profile function $F_q : [1, \infty) \to [0, 1]$, defined as:

$$F_q(t) = \frac{1}{n_p}\text{card}\{p \in P : r_{p,q} \leq t\},$$

where card$\{\cdot\}$ denotes the cardinality of the set. The function $F_q(t)$ represents the cumulative distribution of the performance ratio, and $F_q(1)$ indicates the probability that solver q is the best performer among all considered solvers.

The plotted curves clearly illustrate that the proposed method is quicker and outperforms the other methods on the majority of the test problems. Overall, the numerical results validate the effectiveness and robustness of the proposed approach, which achieves superior performance compared to the other solvers across the majority of the test problems.

We note that the curves for IFR and hDL$^+$ are nearly identical at $t = 1$ because this value is too small to distinguish between the methods. For $t > 1$, hDL$^+$ clearly outperforms IFR and all other methods.

The reason for the good numerical performance is due to the hybrid approach, which generates multiple forms of β_k. These different forms provide greater flexibility to the algorithm, leading to a small number of iterations and less CPU time, thus improving the overall performance.

7. Conclusions

This paper introduced a novel hybrid CG method, hDL$^+$, for training ANNS for early breast cancer prediction and ECG classification as well as large-scale minimization problems. The algorithm satisfies the sufficient descent condition and achieves global convergence under the strong Wolfe line search strategy. The method was successfully applied to train neural networks for early breast cancer prediction, achieving 98.24% accuracy with a precision, recall, and F1-score of 0.99, 0.97, and 0.98, respectively.

In addition, on the ECG classification dataset, the proposed hDL$^+$ achieved an accuracy of 80.45%, demonstrating strong generalization performance across different medical diagnostic tasks.

On a collection of test problems chosen from the CUTE library, the numerical comparisons with CG methods show that hDL$^+$ is reliable and performs better than current CG approaches.

Finally, the main contribution of this work is the hybrid approach that generates multiple forms of β_k, providing greater flexibility and resulting in a small number of iterations and reduced CPU time. These results highlight the effectiveness of the proposed method in both early breast cancer prediction and ECG classification tasks.

Author contributions

Mehamdia Abd Elhamid: Conceptualization, methodology, formal analysis, writing—original draft; Raouf Ziadi: Methodology, data curation, investigation, writing—original draft; Alaa Luqman Ibrahim: Validation, software, investigation; Mohammed A. Saleh: Formal analysis, software, supervision, writing—review & editing; Abdulgader Z. Almaymuni: Project administration, software, funding acquisition, writing—review & editing. All authors reviewed the results and approved the final version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University (www.qu.edu.sa) for financial support (QU-APC-2026).

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. N. Andrei, An unconstrained optimization test functions, *Adv. Model. Optim.*, **10** (2008), 147–161.
2. M. R. Arazm, S. Babaie-Kafaki, R. Ghanbari, An extended Dai-Liao CG method with global convergence for nonconvex functions, *Glasnik Mat.*, **52** (2017), 361–375. https://doi.org/10.3336/gm.52.2.12
3. S. Babaie-Kafaki, On the sufficient descent condition of the Hager-Zhang CG methods, *4OR*, **12** (2014) 285–292. https://doi.org/10.1007/s10288-014-0255-6
4. V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, A. Pedone, A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: The breast cancer classification problem, In: *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, 1958–1965. https://doi.org/10.1109/IJCNN.2006.1716350

5. A. Bir-Jmel, S. M. Douiri, S. EI Bernoussi, A. Maafiri, Y. Himeur, S. Atalla, et al., GFLASSO-LR: Logistic regression with generalized fused LASSO for gene selection in high-dimensional cancer classification, *Computers*, **13** (2024), 93. https://doi.org/10.3390/computers13040093
6. I. Bongartz, A. R. Conn, N. Gould, P. L. Toint, Constrained and unconstrained testing environment, *ACM Trans. Math. Software*, **21** (1995), 123–160. https://doi.org/10.1145/200979.201043
7. Y. Chaib, A. E. Mehamdia, Global convergence of hybrid CG method and its application to nonparametric estimation, *Math. Found. Comput.*, **8** (2025), 296–309. https://doi.org/10.3934/mfc.2024011
8. Y. Chaib, A. E. Mehamdia, Global convergence of new CG methods with application in conditional model regression, *Iran. J. Numer. Anal. Optim.*, **15** (2025), 1–26. https://doi.org/10.22067/ijnao.2024.85389.1344
9. S. B. Cho, H. H. Won, Cancer classification using ensemble of neural networks with multiple significant gene subsets, *Appl. Intell.*, **26** (2007), 243–250. https://doi.org/10.1007/s10489-006-0020-4
10. Y. H. Dai, Y. Yuan, A nonlinear CG method with a strong global convergence property, *SIAM J. Optim.*, **10** (1999), 177–182. https://doi.org/10.1137/S1052623497318992
11. Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, X. Yin, Y. Yuan, Convergence properties of nonlinear conjugate gradient methods, *SIAM J. Optim.*, **10** (2000), 348–358. https://doi.org/10.1137/S1052623494268443
12. Y. H. Dai, C. X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search, *SIAM J. Optim.*, **23** (2013), 296–320. https://doi.org/10.1137/100813026
13. Z. Dai, F. Wen, Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property, *Appl. Math. Comput.*, **218** (2012), 7421–7430. https://doi.org/10.1016/j.amc.2011.12.091
14. Z. Dai, T. Wu, The impact of oil shocks on systemic risk of the commodity markets, *J. Syst. Sci. Complex.*, **37** (2024), 2697–2720. https://doi.org/10.1007/s11424-024-3224-y
15. E. D. Dolan, J. J. Morè, Benchmarking optimization software with performance profiles, *Math. Program.*, **91** (2002), 201–213. https://doi.org/10.1007/s101070100263
16. R. Fletcher, C. Reeves, Function minimization by CGs, *Comput. J.*, **7** (1964), 149–154. https://doi.org/10.1093/comjnl/7.2.149
17. J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Pineros, et al., Global cancer observatory: Cancer today, In: *Lyon: International agency for research on cancer, 2020*, 557–567.
18. Y. Habchi, Y. Himeur, H. Kheddar, A. Boukabou, S. Atalla, A. Chouchane, et al., AI in thyroid cancer diagnosis: Techniques, trends, and future directions, *Systems*, **11** (2023), 519. https://doi.org/10.3390/systems11100519
19. Y. Habchi, H. Kheddar, Y. Himeur, A. Belouchrani, E. Serpedin, F. Khelifi, et al., Advanced deep learning and large language models: Comprehensive insights for cancer detection, *Image Vision Comput.*, **157** (2025), 105495. https://doi.org/10.1016/j.imavis.2025.105495
20. M. R. Hestenes, E. L. Stiefel, Methods of CGs for solving linear systems, *J. Res. Natl. Bur. Stand.*, **49** (1952), 409–436.

21. X. Z. Jiang, J. B. Jian, Two modified nonlinear conjugate gradient methods with disturbance factors for unconstrained optimization, *Nonlinear Dyn.*, **77** (2014), 387–397. https://doi.org/10.1007/s11071-014-1303-7
22. Y. Hu, K. Ashenayi, R. Veltri, G. O'Dowd, G. Miller, R. Hurst, et al., A comparison of neural network and fuzzy c-means methods in bladder cancer cell classification, In: *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, 1994, 3461–3466. https://doi.org/10.1109/ICNN.1994.374891
23. A. L. Ibrahim, B. G. Fathi, M. B. Abdulrazzaq, An improved three-term CG approach for deep neural network training in ECG signal classification, *Netw. Model. Anal. Health Inform. Bioinforma*, **14** (2025), 141. https://doi.org/10.1007/s13721-025-00639-6
24. A. L. Ibrahim, B. G. Fathi, M. B. Abdulrazzaq, Improving three-term CG methods for training artificial neural networks in accurate heart disease prediction, *Neural Comput. Appl.*, **37** (2025), 10381–10405. https://doi.org/10.1007/s00521-025-11121-9
25. A. Maiza, R. Ziadi, M. A. Saleh, A. Z. Almaymuni, An improved conjugate gradient algorithm by adapting a new line search technique, *Electron. Res. Arch.*, **33** (2025), 2148–2171. https://doi.org/10.3934/era.2025094
26. A. Maiza, R. Ziadi, M. A. Saleh, A. Z. Almaymuni, A combination of two conjugate gradient methods under a new line search with its application in image restoration problems, *Int. J. Appl. Math. Comput. Sci*, **35** (2025), 2. https://doi.org/10.61822/amcs-2025-0019
27. K. Mc Namara, H. Alzubaidi, J. K. Jackson, Cardiovascular disease as a leading cause of death: how are pharmacists getting involved? *Integr. Pharm. Res. Pract.*, **8** (2019), 1–11. https://doi.org/10.2147/IPRP.S133088
28. A. E. Mehamdia, Y. Chaib, Improved CG methods and application to nonparametric estimation, *Appl. Math.*, **51** (2024), 147–161. https://doi.org/10.4064/am2512-6-2024
29. A. E. Mehamdia, Y. Chaib, A modified CG method for solving unconstrained optimization with application in conditional mode regression, *Int. J. Comput. Math.*, **102** (2025), 1415–1427. https://doi.org/10.1080/00207160.2025.2499877
30. A. E. Mehamdia, Y. Chaib, T. Bechouat, Two improved nonlinear CG methods with application in conditional model regression function, *J. Ind. Manag. Optim.*, **21** (2025), 658–675. https://doi.org/10.3934/jimo.2024098
31. A. E. Mehamdia, Y. Chaib, Some improved nonlinear CG methods and application to non-parametric estimation, *Filomat*, **39** (2025), 9213–9234. https://doi.org/10.2298/FIL2526213M
32. A. Ramady, H. M. Ahmed, K. A. Ismail, W. B. Rabie, Exploring soliton families in β fractional coupled NLSEs with variable coefficients using improved modified extended tanh-function method, *J. Appl. Anal. Comput.*, **16** (2026) 2264–2286. https://doi.org/10.11948/20250191
33. W. B. Rabie, H. M. Ahmed, A. Abd-Elmonem, N. Alhubieshi, D. I. Elimy, Dispersive solitons and modulational stability in magneto-optic waveguides with β-fractional derivative and Kudryashov's nonlinearity, *Arch. Comput. Methods Eng.*, 2026. https://doi.org/10.1007/s11831-026-10645-0
34. W. B. Rabie, H. M. Ahmed, M. E. Ramadan, N. S. E. Abdalla, A. Abd-Elmonem, T. A. Sulaiman, et al., Dual-soliton structures and stability analysis in a nonlinear fractional Schrödinger equation with dual-mode dispersion, *Mod. Phys. Lett. A*, **41** (2026), 2650077. https://doi.org/10.1142/S021773232650077X

35. A. E. Rateb, H. M. Ahmed, A. Darwish, M. Ammar, W. B. Rabie, Analytical wave families and stability dynamics in a modified complex Ginzburg-Landau model, *Sci. Rep.*, **16** (2026), 7485. https://doi.org/10.1038/s41598-026-37824-0
36. W. B. Rabie, H. B. Amer, H. Khan, J. Alzabut, Exact solutions and stability thresholds for the fractional gardner equation with high-order dispersion, *Eur. J. Pure. Appl. Math.*, **19** (2026), 6805. https://doi.org/10.29020/nybg.ejpam.v19i1.6805
37. H. M. Rai, A. Trivedi, ECG signal classification using wavelet transform and back propagation neural network, In: *2012 5th International Conference on Computers and Devices for Communication (CODEC)*, 2012, 1–4. https://doi.org/10.1109/CODEC.2012.6509183
38. E. Polak, G. Ribiere, Notesurla convergence de directions conjuguees, *Rev. Fr. Inf. Rech. Oper.*, **16** (1969), 35–43.
39. B. T. Polyak, The CG method in extreme problems, *USSR Comput. Math. Math. Phys.*, **9** (1969), 94–112. https://doi.org/10.1016/0041-5553(69)90035-4
40. V. Seena, J. Yomas, A review on feature extraction and denoising of ECG signal using wavelet transform, In: *2014 2nd International Conference on Devices, Circuits and Systems (ICDCS)*, 2014, 1–6. https://doi.org/10.1109/ICDCSyst.2014.6926190
41. H. H. Won, S. B. Cho, Paired neural network with negatively correlated features for cancer classification in DNA gene expression profiles, In: *Proceedings of the International Joint Conference on Neural Networks*, **3** (2003), 1708–1713. https://doi.org/10.1109/IJCNN.2003.1223664
42. R. Xu, X. Cai, D. Wunsch, Gene expression data for DLBCL cancer survival prediction with a combination of machine learning technologies, In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, 894–897. https://doi.org/10.1109/IEMBS.2005.1616559
43. X. Wu, H. Shao, P. Liu, Y. Zhang, Y. Zhuo, An efficient conjugate gradient-based algorithm for unconstrained optimization and its projection extension to large-scale constrained nonlinear equations with applications in signal recovery and image denoising problems, *J. Comput. Appl. Math.*, **422** (2023), 114879. https://doi.org/10.1016/j.cam.2022.114879
44. O. O. Yousif, R. Ziadi, A. Z. Almaymuni, M. A. Saleh, An improved version of Polak-Ribière-Polyak conjugate gradient method with its applications in neural networks training, *Electron. Res. Arch.*, **33** (2025), 4799–4815. https://doi.org/10.3934/era.2025216
45. L. Ziaei, A. R. Mehri, M. Salehi, Application of artificial neural networks in cancer classification and diagnosis prediction of a subtype of lymphoma based on gene expression profile, *JRMS*, **11** (2006), 13–17.
46. G. Zoutendijk, Nonlinear programming computational methods, *Integer Nonlinear Program.*, 1970, 37–86.

AIMS Press