



Linear Dilation-Erosion Perceptron Trained Using a Convex-Concave Procedure

Angélica Lourenço Oliveira and Marcos Eduardo Valle

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 30, 2021

Linear Dilation-Erosion Perceptron Trained Using a Convex-Concave Procedure[★]

Angelica Lourenço Oliveira¹[0000-0002-8689-8522]
and Marcos Eduardo Valle²[0000-0003-4026-5110]

¹ ²Campinas State University, Campinas SP 13083-970, Brazil

¹ ra211686@ime.unicamp.br

² valle@ime.unicamp.br

Abstract. Mathematical morphology (MM) is a theory of non-linear operators used for the processing and analysis of images. Morphological neural networks (MNNs) are neural networks whose neurons compute morphological operators. Dilations and erosions are the elementary operators of MM. From an algebraic point of view, a dilation and an erosion are operators that commute respectively with the supremum and infimum operations. In this paper, we present the *linear dilation-erosion perceptron* (ℓ -DEP), which is given by applying linear transformations before computing a dilation and an erosion. The decision function of the ℓ -DEP model is defined by adding a dilation and an erosion. Furthermore, training a ℓ -DEP can be formulated as a convex-concave optimization problem. We compare the performance of the ℓ -DEP model with other machine learning techniques using several classification problems. The computational experiments support the potential application of the proposed ℓ -DEP model for binary classification tasks.

Keywords: Mathematical morphology · concave-convex optimization · binary classification · continuous piece-wise linear function.

1 Introduction

Machine learning techniques play an important role in both pattern recognition and soft computing. Neural networks, which are inspired by the biological nervous system, are among the most efficient machine learning techniques used to solve pattern recognition tasks including, for example, computer vision and natural language processing [3]. Due to page constraints, we will not provide an extensive review of the many interesting and effective neural network models which have been proposed in the last years but focus on the few models relevant to the development of the linear dilation-erosion perceptron addressed in this paper.

[★] This work was supported in part by the São Paulo Research Foundation (FAPESP) under grant no 2019/02278-2, the National Council for Scientific and Technological Development (CNPq) under grant no 310118/2017-4, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

First of all, it is widely known that the perceptron, introduced by Rosenblatt in the late 1950s, can be used for binary classification tasks [18]. Precisely, the perceptron yields one class if the weighted sum of the inputs is greater than or equal to a threshold and returns the other class otherwise. Furthermore, Rosenblatt proposed a learning rule that converges whenever the samples of the two classes are linearly separable [10]. The multi-layer perceptron (MLP) network, with at least one hidden layer and trained using backpropagation algorithms, overcomes the limitations of the Rosenblatt’s perceptron for non-linearly separable classification tasks [10]. The support vector classifiers (SVCs) developed by Vapnik and collaborators also overcomes the limitations of the perceptrons for binary classification tasks. Like the perceptron, an SVC also predicts the class of a sample by computing a weighted sum. However, training an SVC is performed by maximizing the margin of separation between the two classes [25]. Furthermore, the classification performance of an SVC increases significantly by using the kernel trick. The kernel trick allows an SVC model to be effectively applied to non-linearly classification problems. For comparison purposes, in this paper, we consider both linear and radial-basis function SVC models as well as an MLP network.

In the late 1980s, Ritter and collaborators developed the so-called image algebra as an attempt to provide a unified framework for image processing and analysis techniques [16]. Using the image algebra framework, both linear and morphological operators are defined analogously. By replacing the usual linear dot-product in Rosenblatt’s perceptron model with the corresponding lattice-based operation used to define morphological operators, Ritter and Sussner derived the morphological perceptron in the middle 1990s [14, 15]. The elementary operations from mathematical morphology are dilations and erosions [20]. The maximum of additions yields a dilation while the minimum of sums yields an erosion. Hence, a dilation-based perceptron classifies a pattern by computing the maximum of its components plus the corresponding weights. Dually, the decision function of an erosion-based morphological perceptron is given by the minimum of the pattern’s components plus the weights.

Such as the MLP, a multi-layer morphological perceptron (MLMP) with at least one hidden layer with both dilation-based and erosion-based neurons can theoretically solve any binary classification tasks [17]. From a geometric point of view, an MLMP discriminates two classes by enclosing patterns in hyper-boxes [17, 22]. The non-differentiability of the lattice-based operations, however, makes training morphological neural networks difficult for gradient-based algorithms. Thus, some researchers developed training algorithms in which the network structure grows by adding hyper-boxes to fit the training data [17, 22, 23]. On the downside, growing algorithms usually overfit the training data.

Recently, Charisopoulos and Maragos formulated the training of a morphological perceptron as a concave-convex optimization problem which, in some sense, resembles the training of a linear SVC [4]. Despite the encouraging performance in simple classification tasks, both dilation-based and erosion-based morphological perceptrons implicitly assume a relationship between the input

features and their classes [24]. Using concepts from vector-valued mathematical morphology, one of us proposed the reduced dilation-erosion perceptron (r-DEP), which overcomes the limitations of the morphological perceptron [24]. In a few words, the r-DEP first transforms the input into a feature vector. The feature vector is then classified using a linear combination of a dilation-based and an erosion-based perceptron.

In this paper, a linear transformation is used to map the input to the feature vector. The resulting model is referred to as a *linear dilation-erosion perceptron* (ℓ -DEP). We point out that a ℓ -DEP model can also theoretically solve any binary classification task. Furthermore, inspired by the works of Charisopoulos and Maragos, we formulate the training of a ℓ -DEP classifier as a convex-concave optimization problem.

The paper is organized as follows. The next section provides the mathematical background necessary to understand the models presented in this paper. Section 3 briefly reviews the morphological perceptron as well as the r -DEP model. The ℓ -DEP model and its training algorithm are introduced in Section 4. Computational experiments comparing the performance of the ℓ -DEP model with other classical models from the literature are given in Section 5. The paper finishes with some concluding remarks in Section 6.

2 Basic Concepts on Mathematical Morphology

Mathematical morphology is a non-linear theory widely used for image processing and analysis [12, 20]. From the theoretical point of view, mathematical morphology is very well defined on complete lattices. A partially ordered set (\mathbb{L}, \preceq) is a complete lattice if any subset of \mathbb{L} has both a supremum and an infimum, which are denoted respectively by $\sup X$ and $\inf X$.

The elementary operators from mathematical morphology commute with the lattice-based operations. Formally, given complete lattices \mathbb{L} and \mathbb{M} , a dilation $\delta : \mathbb{L} \rightarrow \mathbb{M}$ and an erosion $\varepsilon : \mathbb{L} \rightarrow \mathbb{M}$ satisfy respectively the following identities for all $X \in \mathbb{L}$ [11]:

$$\delta(\sup X) = \sup_{\mathbf{x} \in X} \{\delta(\mathbf{x})\} \quad \text{and} \quad \varepsilon(\inf X) = \inf_{\mathbf{x} \in X} \{\varepsilon(\mathbf{x})\}. \quad (1)$$

Example 1. Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ denote the extended real-numbers. The cartesian product $\bar{\mathbb{R}}^n$ is a complete lattice with the partial ordering given by $\mathbf{x} \preceq \mathbf{y} \Leftrightarrow x_i \leq y_i, \forall i = 1 : n$. Given $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, the operators $\delta_{\mathbf{a}}, \varepsilon_{\mathbf{b}} : \bar{\mathbb{R}}^n \rightarrow \bar{\mathbb{R}}$ given by

$$\delta_{\mathbf{a}}(\mathbf{x}) = \max_{j=1:n} \{a_j + x_j\} \quad \text{and} \quad \varepsilon_{\mathbf{b}}(\mathbf{x}) = \min_{j=1:n} \{b_j + x_j\}, \quad (2)$$

for all $\mathbf{x} \in \bar{\mathbb{R}}^n$, are respectively a dilation and an erosion [23].

The lattice-based elementary operations of mathematical morphology can be extended to more abstract sets using the concept of reduced orderings [8]. Let

we briefly review the concepts of reduced dilation and reduced erosion as defined recently in [24]. Let \mathbb{V} and \mathbb{W} be an arbitrary sets, not necessarily complete lattices. Also, let $\rho : \mathbb{V} \rightarrow \mathbb{L}$ and $\sigma : \mathbb{W} \rightarrow \mathbb{M}$ be surjective mappings, where \mathbb{L} and \mathbb{M} are both complete lattices. An operator $\delta^r : \mathbb{V} \rightarrow \mathbb{W}$ is a reduced dilation and an operator $\varepsilon^r : \mathbb{V} \rightarrow \mathbb{W}$ is a reduced erosion if there exist a dilation $\delta : \mathbb{L} \rightarrow \mathbb{M}$ and an erosion $\varepsilon : \mathbb{L} \rightarrow \mathbb{M}$ such that

$$\sigma(\delta^r(\mathbf{x})) = \delta(\rho(\mathbf{x})) \quad \text{and} \quad \sigma(\varepsilon^r(\mathbf{x})) = \varepsilon(\rho(\mathbf{x})), \quad \forall \mathbf{x} \in \mathbb{V}. \quad (3)$$

Reduced morphological operators, such as the reduced dilation and the reduced erosion defined above, have been effectively used for processing vector-valued images such as color and hyperspectral images [26, 27].

3 Reduced Dilation-Erosion Perceptron

Morphological neural networks (MNNs) are neural networks whose induced local field of the neurons are given by an operation from mathematical morphology [23]. The morphological perceptrons, introduced by Ritter and Sussner in the middle 1990s for binary classification tasks, are one of the earliest MNNs [14]. In a few words, the morphological perceptrons are obtained by replacing the usual affine transformation $A(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$, for all $\mathbf{x} \in \mathbb{R}^n$, by an elementary morphological operator in Rosenblatt's perceptron given by $y = f(A(\mathbf{x})) \equiv fA(\mathbf{x})$, where f denotes a hard limiter activation function [9]. Formally, the two morphological perceptrons are given by the equations $y = f(\delta_{\mathbf{a}}(\mathbf{x})) \equiv f\delta_{\mathbf{a}}(\mathbf{x})$ and $y = f(\varepsilon_{\mathbf{b}}(\mathbf{x})) \equiv f\varepsilon_{\mathbf{b}}(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{R}^n$. Specifically, the dilation-based perceptron and the erosion-based perceptron are given respectively by

$$y = f\left(\max_{j=1:n}\{a_j + x_j\}\right) \quad \text{or} \quad y = f\left(\min_{j=1:n}\{b_j + x_j\}\right), \quad \mathbf{x} \in \mathbb{R}^n. \quad (4)$$

For simplicity, in this paper we shall consider the signal function as the hard limiter activation function of the morphological perceptrons, i.e., $f(x) = +1$ if $x \geq 0$ and $f(x) = 0$, otherwise. Note that a morphological perceptron model is given by either the composition $f\delta_{\mathbf{a}}$ or the composition $f\varepsilon_{\mathbf{b}}$, where $\delta_{\mathbf{a}} : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\varepsilon_{\mathbf{b}} : \mathbb{R}^n \rightarrow \mathbb{R}$ denote respectively the dilation and the erosion given (1), for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. Because of the maximum operation, a dilation-based perceptron given by $y = f\delta_{\mathbf{a}}$ favors the positive class whose label is +1 [24]. Dually, an erosion-based perceptron defined by $y = f\varepsilon_{\mathbf{b}}$ favors the negative class, i.e., the class label -1, because of the minimum operation.

In [1], Araujo proposed an hybrid MNN called dilation-erosion perceptron (DEP). For binary classification tasks, a DEP computes the convex combination of a dilation and an erosion followed by a hard-limiter function. Intuitively, the DEP model allows for a balance between the two classes. In mathematical terms, given $\beta \in [0, 1]$, a DEP model is defined by $y = f(\tau(\mathbf{x})) \equiv f\tau(\mathbf{x})$, where

$$\tau(\mathbf{x}) = \beta\delta_{\mathbf{a}} + (1 - \beta)\varepsilon_{\mathbf{b}}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad (5)$$

is the decision function of the DEP model.

Despite the successful application of the DEP model for times series prediction [1], the DEP classifiers have a serious drawback: as a lattice-based model, the DEP classifier presupposes a partial ordering on the input space as well as on the set of classes. To overcome this problem, Valle recently proposed the so-called *reduced dilation-erosion perceptron* (r -DEP) using concepts from vector-valued mathematical morphology [24]. An r -DEP classifier is defined as follows: Let \mathbb{V} be the input space and let $\mathbb{C} = \{c_1, c_2\}$ be the set of class labels of a binary classification task. The input space \mathbb{V} is usually a subset of \mathbb{R}^n , but we may consider more abstract input spaces. Also, consider the complete lattice $\mathbb{L} = \bar{\mathbb{R}}^r$, equipped with the usual component-wise ordering, and let us denote $\mathbb{M} = \{-1, +1\}$. Given a one-to-one correspondence $\sigma : \mathbb{C} \rightarrow \mathbb{M}$ and a surjective mapping $\rho : \mathbb{V} \rightarrow \bar{\mathbb{R}}^r$, an r -DEP model is defined by the equation $y = \sigma^{-1}(f(\tau^r(\mathbf{x}))) \equiv \sigma^{-1}f\tau^r(\mathbf{x})$, where $\tau^r : \mathbb{V} \rightarrow \mathbb{R}$ is the decision function given by the following convex combination for $\beta \in [0, 1]$:

$$\tau^r(\mathbf{x}) = \beta\delta_{\mathbf{a}}(\rho(\mathbf{x})) + (1 - \beta)\varepsilon_{\mathbf{b}}(\rho(\mathbf{x})), \quad \forall \mathbf{x} \in \mathbb{V}. \quad (6)$$

Finally, let us briefly comment on the training of an r -DEP classifier. Consider a training set $\mathcal{T} = \{(\mathbf{x}_i, d_i) : i = 1, \dots, m\} \subseteq \mathbb{V} \times \mathbb{C}$. From equation (6), training an r -DEP is done simply training a DEP classifier using the transformed training data $\mathcal{T}^r = \{(\rho(\mathbf{x}_i), \sigma(d_i)) : i = 1, \dots, m\} \subseteq \mathbb{L} \times \mathbb{M}$. At this point, we would like to recall that Charisopoulos and Maragos formulated the training of the morphological perceptrons as well as the hybrid DEP classifier as the solution of a convex-concave programming (CCP) problem [4]. Training r -DEP classifier using convex-concave procedures yielded encouraging results on classification tasks [24].

4 Linear Dilation-Erosion Perceptron

The main challenge in defining a successful r -DEP classifier is how to determine the surjective mapping $\rho : \mathbb{V} \rightarrow \bar{\mathbb{R}}^r$. In this paper, we propose linear transformations as the mapping ρ , resulting in the so-called *linear dilation-erosion perceptron* (ℓ -DEP) classifier. Precisely, a ℓ -DEP model is defined as follows. Let the input space be the n -dimensional vector space $\mathbb{V} = \mathbb{R}^n$ and let $\mathbb{C} = \{c_1, c_2\}$ be the set of class labels of a binary classification task. Also, let $\mathbb{L}_1 = \bar{\mathbb{R}}^{r_1}$ and $\mathbb{L}_2 = \bar{\mathbb{R}}^{r_2}$ be complete lattices with the usual component-wise partial ordering. Given linear mappings $\rho_1 : \mathbb{V} \rightarrow \mathbb{L}_1$ and $\rho_2 : \mathbb{V} \rightarrow \mathbb{L}_2$, the application of linear mappings before the evaluation of an elementary morphological mapping yields the decision function $\tau^\ell : \mathbb{V} \rightarrow \mathbb{R}$ given by

$$\tau^\ell(\mathbf{x}) = \beta\delta_{\mathbf{a}}(\rho_1(\mathbf{x})) + (1 - \beta)\varepsilon_{\mathbf{b}}(\rho_2(\mathbf{x})), \quad \mathbf{x} \in \mathbb{V}.$$

Finally, such as the previous model, the ℓ -DEP model is defined by the equation

$$y = \sigma^{-1}(f(\tau^\ell(\mathbf{x}))) \equiv \sigma^{-1}f\tau^\ell(\mathbf{x}), \quad (7)$$

where f is the sign function and $\sigma : \mathbb{C} \rightarrow \{-1, +1\}$ is a one-to-one mapping.

In contrast to the decision function τ^r of the r -DEP model given by (6), the decision function τ^l of the ℓ -DEP is defined using two linear mappings ρ_1 and ρ_2 . As a consequence, τ^l is a continuous piece-wise linear function which is able to approximate any continuous-valued function from a compact on \mathbb{R}^n to \mathbb{R} [7, 21]. Let us formalize this remark. The linear mappings ρ_1 and ρ_2 satisfies $\rho_1(\mathbf{x}) = R_1\mathbf{x}$ and $\rho_2(\mathbf{x}) = R_2\mathbf{x}$ for real-valued matrices $R_1 \in \mathbb{R}^{r_1 \times n}$ and $R_2 \in \mathbb{R}^{r_2 \times n}$. Defining $W = \beta R_1 \in \mathbb{R}^{r_1 \times n}$, $M = (\beta - 1)R_2 \in \mathbb{R}^{r_2 \times n}$, $\mathbf{c} = \beta\mathbf{a}$, and $\mathbf{d} = (\beta - 1)\mathbf{b}$, the decision function of the ℓ -DEP classifier can be alternatively written as

$$\begin{aligned} \tau^l(\mathbf{x}) &= \beta\delta_{\mathbf{a}}(R_1\mathbf{x}) + (1 - \beta)\varepsilon_{\mathbf{b}}(R_2\mathbf{x}) \\ &= \beta \max\{R_1\mathbf{x} + \mathbf{a}\} - (\beta - 1) \min\{R_2\mathbf{x} + \mathbf{b}\} \\ &= \max\{\beta R_1\mathbf{x} + \beta\mathbf{a}\} - \max\{(\beta - 1)R_2\mathbf{x} + (\beta - 1)\mathbf{b}\} \\ &= \max\{W\mathbf{x} + \mathbf{c}\} - \max\{M\mathbf{x} + \mathbf{d}\} \\ &= \max_{i=1:r_1} \{\mathbf{w}_i^T \mathbf{x} + c_i\} - \max_{j=1:r_2} \{\mathbf{m}_j^T \mathbf{x} + d_j\}, \end{aligned} \quad (8)$$

where \mathbf{w}_i^T and \mathbf{m}_j^T are rows of W and M , respectively. From the last identity, we can identify τ^l with a piece-wise linear function [28]. Moreover, from Theorem 4.3 in [7], τ^l is an universal approximator. Thus, a ℓ -DEP classifier can theoretically solve any binary classification problem. Let us now address the training of a ℓ -DEP classifier.

Given a training dataset $T = \{(\mathbf{x}_k, d_k) : k = 1, \dots, m\} \subseteq \mathbb{V} \times \mathbb{C}$, define the sets $C^+ = \{\mathbf{x}_k : \sigma(d_k) = +1\}$ and $C^- = \{\mathbf{x}_k : \sigma(d_k) = -1\}$ of input samples. Inspired by the linear SVC and the concave-convex procedure developed by Charisopoulos and Maragos [4, 10], the parameters W , \mathbf{c} , M and \mathbf{d} of a ℓ -DEP classifier can be determined by minimizing the hinge loss function subject to the constraints

$$\tau^l(\mathbf{x}_k) \leq -1 + \xi_k, \text{ if } \mathbf{x}_k \in C^-, \quad \text{and} \quad \tau^l(\mathbf{x}_k) \geq 1 - \xi_k, \text{ if } \mathbf{x}_k \in C^+. \quad (9)$$

The terms -1 and $+1$ in the left-hand side of the inequalities imposes a margin of separation between the two classes. The slack variables ξ_k , for $k = 1, \dots, m$, allows some classification errors in the training set. Note that, regardless of its class, the input sample \mathbf{x}_k is well classified if $\xi_k \leq 0$. The hinge loss minimizes the positive values of ξ_k , for $\xi = 1, \dots, m$.

Concluding, from (8), a ℓ -DEP classifier is trained by solving the following disciplined convex-concave programming (DCCP) problem [29]:

$$\left\{ \begin{array}{l} \text{minimize}_{W, \mathbf{c}, M, \mathbf{d}, \xi} \quad \sum_{k=1}^m \max(\xi_k, 0), \\ \text{s.t.} \quad \max_{i=1:r_1} (\mathbf{w}_i^T \mathbf{x}_k + c_i) + 1 \leq \max_{j=1:r_2} (\mathbf{m}_j^T \mathbf{x}_k + d_j) + \xi_k, \forall \mathbf{x}_k \in C^- \\ \quad \max_{i=1:r_1} (\mathbf{w}_i^T \mathbf{x}_k + c_i) + \xi_k \geq \max_{j=1:r_2} (\mathbf{m}_j^T \mathbf{x}_k + d_j) + 1, \forall \mathbf{x}_k \in C^+ \end{array} \right. \quad (10)$$

We would like to point out that we solved the disciplined convex-concave programming problem (10) using the algorithm proposed by Shen et. al. [19]. Precisely, in our computational implementations, we solved (10) using the CVXPY

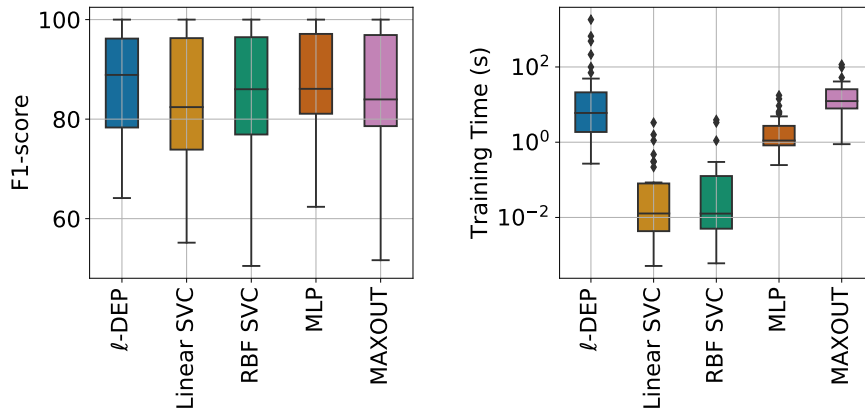


Fig. 1. Boxplot of the F1-score and time required to train the binary classifiers

package [5], which has an extension for solving DCCP problems, combined with the MOSEK solver [2].

5 Computational Experiments

Let us briefly evaluate the performance of the proposed ℓ -DEP classifier on several datasets from the OpenML repository available at <https://www.openml.org/>. For simplicity, we fixed the parameters $r_1 = r_2 = 10$ of the ℓ -DEP classifier for all datasets. Let us also compare the proposed classifier with a multi-layer perceptron (MLP), the maxout network, and the linear and the RBF SVCs [6, 7, 9]. We used python’s `scikit-learn` (`sklearn`) implementations of the MLP and SVC classifiers with their default parameters [13]. The maxout network has been implemented using `tensorflow` with extra functionalities, that is, the maxout layer of `tensorflow-addons`. In our experiments, we used 10 locally affine regions in the maxout network for all datasets [7].

We would like to point out that we handled missing data using `sklearn`’s `SimpleImputer()` command. Furthermore, we partitioned the data set into training and test sets using the `sklearn`’s `StratifiedKFold()` command with $k = 5$. Finally, since some datasets are not balanced, we used the F1-score to measure the performance of the classifiers.

Table 1 contains the mean and the standard deviation of the F1-score obtained from the five classifiers using stratified 5-fold cross-validation. The boxplot shown in Figure 5 summarizes the scores depicted on this table as well as the execution time for training the classifiers. Although the MLP yielded the largest average performance, the ℓ -DEP model produced the largest median performance. Furthermore, it is clear from the boxplot on the left of Figure 5 that the ℓ -DEP is comparable to the other classifiers. In particular, ℓ -DEP achieved

Table 1. Average and standard deviation of the F1-Score.

	ℓ -DEP	Linear SVC	RBF SVC	MLP	MAXOUT
Accute Inflammations	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0
Australian	83.3 ± 1.9	85.2 ± 3.6	86.1 ± 2.9	86.1 ± 1.9	85.5 ± 3.1
Banana	89.2 ± 3.1	55.2 ± 0.0	90.4 ± 1.1	90.1 ± 1.4	84.0 ± 10.4
Banknote	100 ± 0.0	98.5 ± 0.5	100 ± 0.0	100 ± 0.0	99.5 ± 1.0
Blood Transfusion	77.7 ± 2.4	76.1 ± 0.5	76.6 ± 1.1	79.8 ± 1.5	79.5 ± 2.0
Breast Cancer Wisconsin	96.7 ± 1.1	97.2 ± 0.7	97.7 ± 0.5	97.5 ± 1.4	97.2 ± 0.7
Chess	99.3 ± 0.4	96.8 ± 0.6	98.3 ± 0.7	99.2 ± 0.4	99.1 ± 0.1
Colic	81.5 ± 4.0	81.3 ± 6.6	85.1 ± 3.7	84.8 ± 3.2	81.6 ± 5.9
Credit Approval	80.4 ± 3.3	85.5 ± 2.8	85.7 ± 3.0	84.9 ± 0.9	83.9 ± 2.8
Credit-g	68.4 ± 5.0	75.8 ± 3.0	76.5 ± 4.4	73.8 ± 3.7	68.5 ± 4.6
Cylinder Bands	76.7 ± 5.9	69.4 ± 2.4	77.2 ± 2.8	77.4 ± 5.8	72.8 ± 4.7
Diabetes	70.8 ± 3.6	76.4 ± 3.4	75.2 ± 4.7	76.7 ± 3.7	75.1 ± 3.3
EEG-Eye-State	89.3 ± 1.5	61.7 ± 1.9	68.1 ± 5.0	85.6 ± 3.1	74.1 ± 7.5
Haberman	70.3 ± 3.8	73.2 ± 0.8	73.5 ± 2.6	74.5 ± 2.4	74.2 ± 1.8
Hill-Valley	93.6 ± 1.3	60.6 ± 1.9	50.5 ± 1.2	62.4 ± 2.7	51.7 ± 2.0
lpld	64.1 ± 4.8	71.4 ± 0.4	71.2 ± 0.3	73.1 ± 3.0	70.2 ± 3.1
Internet Advertisements	93.7 ± 1.8	96.0 ± 0.7	96.6 ± 0.3	97.2 ± 0.4	97.0 ± 0.6
Ionosphere	88.9 ± 4.5	88.6 ± 2.7	94.3 ± 2.3	92.6 ± 3.1	91.7 ± 4.5
MOPN-3-7-10	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0
Monks-2	88.0 ± 3.7	65.7 ± 0.2	72.7 ± 2.3	82.4 ± 3.1	83.5 ± 3.2
Mushroom	100 ± 0.0	97.8 ± 0.9	100 ± 0.0	100 ± 0.0	100 ± 0.0
Phoneme	84.9 ± 1.2	77.3 ± 1.6	84.5 ± 1.5	85.6 ± 1.4	81.5 ± 3.5
Pishing Websites	95.7 ± 0.6	90.5 ± 0.7	95.0 ± 0.9	96.9 ± 0.5	95.6 ± 0.6
Sick	97.0 ± 0.7	96.6 ± 1.0	96.3 ± 0.4	97.1 ± 0.7	96.8 ± 1.0
Sonar	77.4 ± 3.5	74.5 ± 3.9	84.2 ± 4.2	84.2 ± 3.5	80.8 ± 4.4
Spambase	92.8 ± 1.5	92.9 ± 0.5	93.3 ± 0.5	94.5 ± 0.9	93.5 ± 0.6
Steel Plates Fault	100 ± 0.0	100 ± 0.0	99.6 ± 0.5	99.9 ± 0.1	99.9 ± 0.1
Thoracic Surgery	76.8 ± 3.1	85.1 ± 0.0	85.1 ± 0.0	83.6 ± 1.8	81.9 ± 2.7
Tic-Tac-Toe	92.6 ± 5.0	65.3 ± 0.2	88.8 ± 2.1	86.3 ± 3.2	95.3 ± 4.5
Titanic	79.0 ± 2.3	77.6 ± 2.0	77.8 ± 1.9	77.7 ± 1.9	77.6 ± 1.7
MEAN ± STD	86.9 ± 10.6	82.4 ± 13.3	86.0 ± 12.0	87.5 ± 10.0	85.7 ± 12.0
MEDIAN ± MAD	88.9 ± 9.0	82.4 ± 11.3	86.1 ± 9.4	86.1 ± 8.3	84.0 ± 9.8

slightly higher performance than the maxout network. As to the training execution time, the linear and the RBF SVCs are the fastest models. Despite its longer, the ℓ -DEP is not quite different from the MLP and maxout models in training time.

6 Concluding Remarks

In this paper, we introduced the linear dilation-erosion perceptron (ℓ -DEP) classifier which is given by a convex combination of the composition of linear transformations and the two elementary morphological operation. Specifically, given a one-to-one mapping σ from the set of class labels \mathbb{C} to $\{+1, -1\}$, a ℓ -DEP classifier is defined by means of the equation $y = \sigma^{-1} f \tau^l(\mathbf{x})$, where the decision function $\tau^l : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by (7). Alternatively, τ^l can be expressed by means of (8) for matrices $W \in \mathbb{R}^{r_1 \times n}$ and $M \in \mathbb{R}^{r_2 \times n}$ and vectors $\mathbf{c} \in \mathbb{R}^{r_1}$ and $\mathbf{d} \in \mathbb{R}^{r_2}$. Except for the one-to-one mapping $\sigma : \mathbb{C} \rightarrow \{-1, +1\}$, which can be

defined arbitrarily, the other parameters W , M , \mathbf{c} , and \mathbf{d} are self-adjusted in the training of the ℓ -DEP model. Moreover, the classifier is trained by solving the disciplined convex-concave programming problem given by (10), which has been inspired by the works of Charisopoulos and Maragos [4]. From a theoretical point of view, the decision function τ^l of the ℓ -DEP is a continuous piece-wise linear function [28]. As a consequence, a ℓ -DEP model can in principle solve any binary classification task [7]. Computational experiments with 30 binary classification problems revealed comparable performance of the proposed ℓ -DEP model with other classifiers from the literature, namely, the linear and RFB SVCs, MLP, and the maxout network.

In the future, we plan to investigate further in detail the concave-convex programming problem used to train a ℓ -DEP classifier. In particular, we intend to include regularization terms in the objective function to improve the generalization capability of the ℓ -DEP classifier.

References

1. de A. Araújo, R.: A class of hybrid morphological perceptrons with application in time series forecasting. *Knowledge-Based Systems* **24**(4), 513–529 (2011). <https://doi.org/10.1016/j.knosys.2011.01.001>
2. ApS, M.: MOSEK Optimizer API for Python. - (2020), <https://docs.mosek.com/9.2/pythonapi.pdf>
3. Aurelien Géron: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O Reilly, Sebastopol, California, USA., 2nd edn. (10 2019)
4. Charisopoulos, V., Maragos, P.: Morphological Perceptrons: Geometry and Training Algorithms. In: *Math. Morph. and Its Appli. to Signal and Image Proc.* pp. 3–15. Springer (2017). https://doi.org/0.1007/978-3-319-57240-6_1
5. Diamond, S., Boyd, S.: CVXPY: A Python-Embedded Modeling Language for Convex Optimization. - (Jun 2016), <https://dl.acm.org/doi/10.5555/2946645.3007036>
6. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016)
7. Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout Networks. In: *International Conference on Machine Learning*. pp. 1319–1327 (2013)
8. Goutsias, J., Heijmans, H.J.A.M., Sivakumar, K.: Morphological Operators for Image Sequences. *Computer vision and image understanding* **62**, 326–346 (1995)
9. Haykin, S.: *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ (1999)
10. Haykin, S.: *Neural Networks and Learning Machines*. Prentice-Hall, Upper Saddle River, NJ, 3rd edition edn. (2009)
11. Heijmans, H.J.A.M., Ronse, C.: The algebraic basis of mathematical morphology I. Dilations and erosions. *Computer Vision, Graphics, and Image Processing* **50**(3), 245–295 (1990). [https://doi.org/10.1016/0734-189X\(90\)90148-O](https://doi.org/10.1016/0734-189X(90)90148-O)
12. Heijmans, H.J.A.M.: *Mathematical Morphology: A Modern Approach in Image Processing Based on Algebra and Geometry*. *SIAM Review* **37**(1), 1–36 (1995)
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: *Scikit-learn: Machine Learning in Python*. *The Journal of Machine Learning Research* **12**(null), 2825–2830 (Nov 2011)

14. Ritter, G.X., Sussner, P.: An Introduction to Morphological Neural Networks. In: Proceedings of the 13th International Conference on Pattern Recognition. pp. 709–717. Vienna, Austria (1996)
15. Ritter, G.X., Sussner, P.: Morphological Perceptrons. In: ISAS'97, Intelligent Systems and Semiotics. Gaithersburg, Maryland (1997)
16. Ritter, G.X., Wilson, J.N., Davidson, J.L.: Image Algebra: An Overview. *Computer Vision, Graphics, and Image Processing* **49**(3), 297–331 (1990)
17. Ritter, G.X., Urcid, G.: Lattice Algebra Approach to Single-Neuron Computation. *IEEE Transactions on Neural Networks* **14**(2), 282–295 (2003)
18. Rosenblatt, F.: The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* **65**, 386–408 (1958)
19. Shen, X., Diamond, S., Gu, Y., Boyd, S.: Disciplined convex-concave programming. In: 2016 IEEE 55th Conference on Decision and Control (CDC). pp. 1009–1014 (2016). <https://doi.org/10.1109/CDC.2016.7798400>
20. Soille, P.: *Morphological Image Analysis*. Springer Verlag, Berlin (1999)
21. Stone, M.H.: The Generalized Weierstrass Approximation Theorem. *Mathematics Magazine* **21**(4), 167–184 (1948). <https://doi.org/10.2307/3029750>, <https://www.jstor.org/stable/3029750>, publisher: Mathematical Association of America
22. Sussner, P.: Morphological perceptron learning. In: IEEE International Symposium on Intelligent Control - Proceedings. pp. 477–482. IEEE (1998). <https://doi.org/10.1109/isc.1998.713708>
23. Sussner, P., Esmi, E.L.: Morphological Perceptrons with Competitive Learning: Lattice-Theoretical Framework and Constructive Learning Algorithm. *Information Sciences* **181**(10), 1929–1950 (2011). <https://doi.org/10.1016/j.ins.2010.03.016>
24. Valle, M.E.: Reduced Dilation-Erosion Perceptron for Binary Classification. *Mathematics* **8**(4), 512 (2020). <https://doi.org/10.3390/math8040512>
25. Vapnik, V.N.: *Statistical Learning Theory*. John Wiley and Sons, New York, NY, USA (1998)
26. Velasco-Forero, S., Angulo, J.: Supervised Ordering in \mathbb{R}^p : Application to Morphological Processing of Hyperspectral Images. *IEEE Transactions on Image Processing* **20**(11), 3301–3308 (Nov 2011). <https://doi.org/10.1109/TIP.2011.2144611>
27. Velasco-Forero, S., Angulo, J.: Vector Ordering and Multispectral Morphological Image Processing. In: Celebi, M.E., Smolka, B. (eds.) *Advances in Low-Level Color Image Processing*. pp. 223–239. Springer Netherlands, Dordrecht (2014). https://doi.org/10.1007/978-94-007-7584-8_7
28. Wang, S.: General constructive representations for continuous piecewise-linear functions. *IEEE Transactions on Circuits and Systems I: Regular Papers* **51**(9), 1889–1896 (2004). <https://doi.org/10.1109/TCSI.2004.834521>
29. Yuille, A.L., Rangarajan, A.: The concave-convex procedure. *Neural Computation* **15**(4), 915–936 (2003). <https://doi.org/10.1162/08997660360581958>