Position Paper: Deep Learning in Hilbert Lattices

Vassilis G. Kaburlasos

Department of Informatics, Computer and Telecommunications Engineering
International Hellenic University (IHU)
Serres Campus, 62124 Greece
Email: vgkabs@ihu.gr

***Abstract*—This work proposes two necessary and sufficient conditions for neural computing namely, first, a Hilbert space H and, second, non-linear transformations in H. The conventional deep learning (neural computing) paradigm is extended hierarchically from the Euclidean (Hilbert) space $\mathbb{R}^N$ to the Hilbert space $\mathbb{G}_1^N$, where $\mathbb{G}_1$ is a *quotient* of *equivalence classes* of *Generalized Intervals' Numbers* (*GINs*). The convex cone $\mathbb{F}_1$ of $\mathbb{G}_1$ is considered, where an element of $\mathbb{F}_1$, namely *Intervals' Number* (*IN*), is an (information) granule that may represent a probability/possibility distribution including real numbers; hence, $\mathbb{G}_1 \supset \mathbb{F}_1 \supset \mathbb{R}$. It turns out that $(\mathbb{F}_1, \preceq)$ is a sublattice of lattice $(\mathbb{G}_1, \preceq)$, where partial order represents semantics; moreover, axiomatic logic can be applied in $\mathbb{F}_1$. Because all the required mathematical instruments are available, neural computing can be pursued in $\mathbb{F}_1^N$ including logic/reasoning toward eXplainable Artificial Intelligence (XAI). The far-reaching potential of proposed tools in practical applications is discussed.**

***Index Terms*—Axiomatic Logic, Big Data, Deep Learning, Hilbert Lattice, Intervals' Number (IN), Neural Computing**

I. Introduction

Conventional modeling, initiated by Isaac Newton regarding Physics [1], was later extended to alternative scientific disciplines including Engineering, Economy, Medicine and other. In recent years, conventional modeling has been extended to Artificial Intelligence (AI) deep learning models [2] including Convolutional Neural Networks (CNNs) for image recognition [3] and Large Language Models (LLMs) for human-like language generation [4].

Humans are subjects of intelligence, whereas the physical world is not. Therefore, modeling (human) intelligence by the same instruments and/or practices that successfully model the physical world, e.g. energy optimization practices, misses critical features of intelligence such as logic/reasoning. Furthermore, humans engage non-numeric percepts such as the degree of truth of propositions, data structures, symbols and other. Historically, the first non-numeric percepts studied effectively have been the truth values of propositions resulting in Boolean algebra [5] and, later, fuzzy sets [6]. A further study of Boolean algebras has ushered to the mathematical Order Theory, or Lattice Theory [7]. The interest in applications of mathematical lattices was sustained [8]. In conclusion, the *Lattice Computing* (LC) *paradigm* has been proposed [9] as a modeling paradigm shift to a lattice data domain, including $\mathbb{R}^N$, where partial order represents semantics [10].

This work has been supported by IHU's Research Committee under project no. 81960 entitled "Artificial Intelligence and Applications".

It has been acknowledged that the closest "mathematical relative" to the Euclidean space is a Hilbert space [11]. As the utility of the Euclidean space has been confirmed meticulously /widely /repetitively during millennia of practice, it is worthwhile to further investigate the potential of an enhanced Hilbert space in AI.

From an "ontological" viewpoint, deep learning models are computational intelligence models that, during training, pursue an optimal approximation of a parametric function $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$ [12]. Better performance is typically reported for models with a larger number of tunable parameters [13] − Note that, lately, the number of parameters of a deep learning model has exceeded 1 trillion [14]. The deep learning neural architecture proposed in this work has the potential to implement far more functions than conventional deep learning models.

Conventional deep learning models are typically developed in the Euclidean (Hilbert) space $\mathbb{R}^N$ [2]. Their success has been attributed to the capacity of digital computer hardware to process vast data fast rather than to compute differently [15]. This work proposes an enhanced deep learning that computes differently as detailed below.

Three conditions for neural computing have been proposed in [15], namely (p1) a Hilbert space H, (p2) non-linear operations in H, and (p3) a derivative in H. Using mathematical results presented below, this work proposes two necessary and sufficient conditions for neural computing including, first, a Hilbert H and, second, non-linear transformations in H.

Lately, a novel Hilbert space has been induced hierarchically from $\mathbb{R}$, namely Hilbert space of *Generalized Intervals' Numbers* (*GINs*), symbolically $\mathbb{G}_1$ [15]. The interest is in the convex cone $\mathbb{F}_1 \subset \mathbb{G}_1$ of *(Type-1) Intervals' Numbers* (*INs*). An IN is an *(information) granule* i.e. a mathematical object [16]–[19], that may represent a probability/possibility distribution including real numbers. INs have already been applied to neural/fuzzy systems [16]–[21]. An IN may represent *all-order data statistics*; hence, it may represent countably infinite real numbers, with arbitrarily small error, using a finite number of real numbers. The latter implies a capacity for dealing with big data, with small memory requirements.

The layout of this paper is as follows. Section II outlines the mathematical background. Section III presents the proposed, IN-based neural network. Section IV demonstrates some computational results. Finally, section V concludes by discussing both our contribution and potential future work.

II. The Mathematical Background

This section summarizes a number of mathematical results [7], [15], [22], [23]. More specifically, it considers *complete* mathematical lattices $(\mathbb{L}, \sqsubseteq)$ with *minimum* and *maximum* elements denoted by o and i, respectively. A *positive valuation* (real) function $v : \mathbb{L} \to \mathbb{R}$ is assumed, which, by definition, satisfies: 1) $v(x) + v(y) = v(x \sqcap y) + v(x \sqcup y)$ and 2) $x \sqsubset y \Rightarrow v(x) < v(y)$. A positive valuation results in two useful axiomatically defined functions, first, a *metric distance* function $d : \mathbb{L}\times\mathbb{L} \to \mathbb{R}_0^+$ given by $d(x, y) = v(x\sqcup y)-v(x\sqcap y)$ and, second, an *order measure* function $\sigma : \mathbb{L} \times \mathbb{L} \to [0, 1]$ given by either function "sigma join" $\sigma_\sqcup(x, u) = v(u)/v(x \sqcup u)$ or "sigma meet" $\sigma_\sqcap(x, u) = v(x \sqcap u)/v(x)$.

A. Fuzzy Lattice Reasoning (FLR)

Any employment of an order measure function $\sigma(.,.)$ is called *fuzzy lattice reasoning*, or *FLR* for short [12], [13], [18], [19].

Semantics is intrinsically represented in a lattice by the partial order relation [10]. Note that the "sigma meet" order measure $\sigma_\sqcap(.,.)$ is a generalization of the "Rule of Bayes". Moreover, both $\sigma_\sqcap(.,.)$ and $\sigma_\sqcup(.,.)$ are widely (though implicitly) used in Fuzzy Inference Systems (FISs) [18].

An order measure and/or a metric distance can be extended to the lattice of intervals in a lattice $(\mathbb{L}, \sqsubseteq)$ based on a *dual isomorphic* function $\theta : \mathbb{L} \to \mathbb{L}$ resulting in the positive valuation function $V([a, b]) = v(\theta(a)) + v(b)$ [15]. The set of intervals is a lattice denoted by $(\mathbb{I}_1 = \mathbb{I} \cup \{\varnothing\}, \subseteq)$, where $\mathbb{I}$ is the set of conventional intervals, and $\subseteq$ is the conventional set inclusion relation. The empty set $\varnothing$ in the complete lattice $(\mathbb{I}_1, \subseteq)$ is represented as $\varnothing = [i, o]$.

A lattice $(\mathbb{L}, \sqsubseteq)$ may be the Cartesian product of a set of "constituent" lattices $\mathbb{L}_i$, $i \in I$, where I is in an index set.

B. The Hilbert Lattice of Generalized Intervals' Numbers

Following a set-theoretic approach, this work considers an index set I, regarding constituent lattices $\mathbb{L}_i, i \in I$, of uncountably infinite cardinality $\aleph_1$. In particular, the interest focuses on the complete, totally ordered lattice $(\overline{\mathbb{R}}, \leq)$, where $\overline{\mathbb{R}} = \mathbb{R}\cup\{-\infty, \infty\}$, $\mathbb{R}$ is the set of real numbers, with $o = -\infty$ and $i = +\infty$. The corresponding infimum and supremum operators are denoted by $\wedge$ and $\vee$, respectively.

It is known that the lattice $(V_\mathbb{G}, \sqsubseteq) = (\overline{\mathbb{R}} \times \overline{\mathbb{R}}, \geq \times \leq)$ of *generalized intervals* is a *vector space* over $\mathbb{R}$, where the *addition* is defined as $[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$, moreover the *scalar multiplication* is defined as $\lambda[a, b] = [\lambda a, \lambda b]$ [12].

Lately, it has been shown that the mapping $<,>: V_\mathbb{G}\times V_\mathbb{G} \to \mathbb{R}$ given by $< [a_1, b_1], [a_2, b_2] >= \frac{1}{2}(a_1a_2 + b_1b_2)$ is an inner product in the vector lattice $(V_\mathbb{G}, \sqsubseteq)$. Hence, lattice $(V_\mathbb{G}, \sqsubseteq)$ is a *pre-Hilbert space* with *norm* $\| [a, b] \|= \sqrt{\frac{1}{2}(a^2 + b^2)}$. It turns out that lattice $(V_\mathbb{G}, \sqsubseteq)$ is a *Hilbert space* because every Cauchy sequence in $V_\mathbb{G}$ converges in $V_\mathbb{G}$ with respect to the metric distance $\| [a_1, b_1] - [a_2, b_2] \|$ [15].

Consider the Cartesian product lattice $(\mathbb{G}, \preceq) = \times(V_\mathbb{G}, \sqsubseteq)_h$, where $h \in [0, 1]$. An element of $(\mathbb{G}, \preceq)$, namely *Generalized Intervals' Number* (*GIN*), is a function $E : [0, 1] \to (V_\mathbb{G}, \sqsubseteq) = (\overline{\mathbb{R}} \times \overline{\mathbb{R}}, \geq \times \leq)$. A value $E(h)$ is, alternatively, denoted as E_h. The lattice $(\mathbb{G}, \preceq)$ is a vector space over $\mathbb{R}$, where addition is defined as $E_s(h) = E_1(h) + E_2(h)$, $h \in [0, 1]$ and scalar multiplication is defined as $E_p(h) = \lambda E(h)$, $h \in [0, 1]$.

Next, the interest focuses on the vector space $\mathbb{G}_1$ defined as $\mathbb{G}_1 = \{[a(h), b(h)] : a(.), b(.) \in L_2\}$, where L_2 is the Lebesgue space of square-integrable (real) functions [15]. Specifically, $\mathbb{G}_1$ is a *quotient* space whose elements are the equivalence classes of $\mathbb{G}$ defined such that two INs in an equivalence class of $\mathbb{G}$ differ on a null set. It turns out that the mapping $<,>: \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{R}$ given by

$$< E, F >= \int_0^1 < [a_1(h), b_1(h)], [a_2(h), b_2(h)] > dh, \quad (1)$$

is an inner product in vector space $\mathbb{G}_1$. Hence, the set $\mathbb{G}_1$ of equivalence classes in $\mathbb{G}$ is a *pre-Hilbert space* with norm $\| F \|= \sqrt{\frac{1}{2}\int_0^1 (a_h^2 + b_h^2)dh}$, where $F = [a_h, b_h]$, $h \in [0, 1]$. A potentially useful result is the Cauchy-Schwarz inequality $| < v, w > | \leq \| v \| \| w \|$, which introduces the *similarity cosine* function $\rho : V \times V \to [-1, 1]$ as

$$\rho(v, w) = \frac{< v, w >}{\| v \| \| w \|} \quad (2)$$

toward calculating the angle $\angle(v, w)$ between two vectors $v, w \in V$ as $\angle(v, w) = arccos\frac{<v,w>}{\|v\|\|w\|}$.

It turns out that $\mathbb{G}_1$ is a *Hilbert space* because every Cauchy sequence in $\mathbb{G}_1$ converges in $\mathbb{G}_1$ with respect to the metric $\| F - E \|= \sqrt{\frac{1}{2}\int_0^1 [(a_1(h) - a_2(h))^2 + (b_1(h) - b_2(h))^2]dh}$, where $F = [a_1(h), b_1(h)]$ and $E = [a_2(h), b_2(h)]$, $h \in [0, 1]$. Moreover, the Hilbert space $\mathbb{G}_1$ is lattice ordered [15]. In conclusion, $(\mathbb{G}_1, \preceq)$ is called *Hilbert lattice*.

C. The Convex Cone Sublattice of Intervals' Numbers (INs)

A (Type-1) *Intervals' Number*, or *IN* for short, is defined as a function $E : [0, 1] \to \mathbb{I}_1$, which satisfies: (C1) $h_1 \leq h_2 \Rightarrow E_{h_1} \supseteq E_{h_2}$ and (C2) $\forall S \subseteq [0, 1] : \bigcap_{h\in S} E_h = E_{\vee S}$. The family of INs is denoted by $\mathbb{F}$. It turns out that $(\mathbb{F}, \preceq)$ is a complete lattice, where $\preceq$ is defined as

$$P \preceq Q \Leftrightarrow P_h \subseteq Q_h, \Leftrightarrow P(x) \leq Q(x), \quad (3)$$

where $h \in [0, 1]$ and $x \in \mathbb{R}$.

Probability and possibility distributions have been studied comparatively [24], [25]. An IN, previously called FIN (Fuzzy Intervals' Number) [26], is a mathematical object which may represent either a possibility distribution [18], [26] or a probability distribution [16]. In the following, the interest focuses on the quotient space $\mathbb{F}_1$ whose elements are equivalence classes in $\mathbb{F}$. From here forward, an element of $\mathbb{F}_1$ will be called (Type-1) IN.

Fig.1 displays INs with interval support $[1.90, 3.10]$. Each IN was induced from 62 samples. In particular, IN P in Fig.1(a) was induced by Algorithm "distrIN" [27], whereas IN Q in Fig.1(b) was induced by Algorithm "CALFIN" [23]. Fig.1(c) and Fig.1(d) display the corresponding "interval representations" of the equivalent "membership-function representations" of the INs P and Q shown Fig.1(a) and Fig.1(b), respectively. Note that the interval representation of an IN requires two numbers per level i.e. the two interval ends. In particular, regarding the interval representation of IN P in Fig.1(c) only one number is required per level because all the right interval ends coincide. In conclusion, the IN P in Fig.1(c) is represented by $L = 32$ numbers ordered increasingly.

The IN P in Fig.1(a) corresponds to a Cumulative Distribution Function (CDF) induced from a population of 62 numerical data samples indicated with an "x" mark; it is $P(2.46) = 0.5$. Actually, the IN P in Fig.1(a) represents only part of a CDF since a CDF in non-decreasing on its domain, whereas function $P(x)$ drops to 0 as soon as $P(x)$ reaches its largest value of 1. Nevertheless, apparently, there is a bijective mapping between the proper subset $S = \{F \in \mathbb{F}_1$: $(F_0 = [a_0, b_0]$ with $-\infty < a_0 \leq b_0 < +\infty)$.AND.$(a_0 \leq x_1 \leq x_2 \leq b_0 \Rightarrow F(x_1) \leq F(x_2))\}$ of $\mathbb{F}_1$ and CDFs; in the aforementioned sense, an IN represents a CDF.

An IN represents an *information granule*. An advantage of an IN is its potential to represent countably infinite numbers i.e. *all-order data statistics*, potentially to any degree of accuracy depending on the number L of levels [18]. An interval $F_h \in \mathbb{I}_1, h \in [0, 1]$ is denoted either by $[a_h, b_h]$ or, alternatively, by $[a(h), b(h)], h \in [0, 1]$. The set of INs is lattice-ordered according to (3). The corresponding lattice of INs is denoted by $(\mathbb{F}_1, \preceq)$ and it is a sublattice of lattice $(\mathbb{G}_1, \preceq)$. In particular, the set of *trivial* INs $[x, x]_h, x \in \mathbb{R}, h \in [0, 1]$ corresponds to the Hilbert space $\mathbb{R}$ of real numbers. However, $\mathbb{F}_1$ is not a vector space because for a non-trivial $P \in \mathbb{F}_1$ it follows $(-P) \notin \mathbb{F}_1$. It turns out that $\mathbb{F}_1$ is a *convex cone* in $\mathbb{G}_1$ because if $P, Q \in \mathbb{F}_1$ then $(\lambda P + \mu Q) \in \mathbb{F}_1, \forall \lambda, \mu \geq 0$. In conclusion, $\mathbb{F}_1$ is a lattice-ordered convex cone. Moreover, the set $\mathbb{F}_1$ is *complete* in the sense that every Cauchy sequence in $\mathbb{F}_1$ converges in $\mathbb{F}_1$ [15].

Given $E, F \in \mathbb{F}_1$, there follow a *metric distance* function $D_1 : \mathbb{F}_1 \times \mathbb{F}_1 \to \mathbb{R}_0^+$, two *order measure* functions $\sigma : \mathbb{F}_1 \times \mathbb{F}_1 \to [0, 1]$ and a *similarity cosine* function $\rho_1 : \mathbb{F}_1 \times \mathbb{F}_1 \to [-1, 1]$, respectively, as

$$D_1(E, F; v, \theta) = \int_0^1 d_1(E_h, F_h; v, \theta) dh, \tag{4}$$

where $d_1 : \mathbb{I}_1 \times \mathbb{I}_1 \to \mathbb{R}_0^+$ is a *metric distance* function given by $d_1([a_1, b_1], [a_2, b_2]; v, \theta) = [v(\theta(a_1 \wedge a_2)) - v(\theta(a_1 \vee a_2))] + [v(b_1 \vee b_2) - v(b_1 \wedge b_2)]$;

$$\sigma_{\curlywedge}(E, F; v, \theta) = \int_0^1 \sigma_{\sqcap}(E_h, F_h; v, \theta) dh, \tag{5}$$

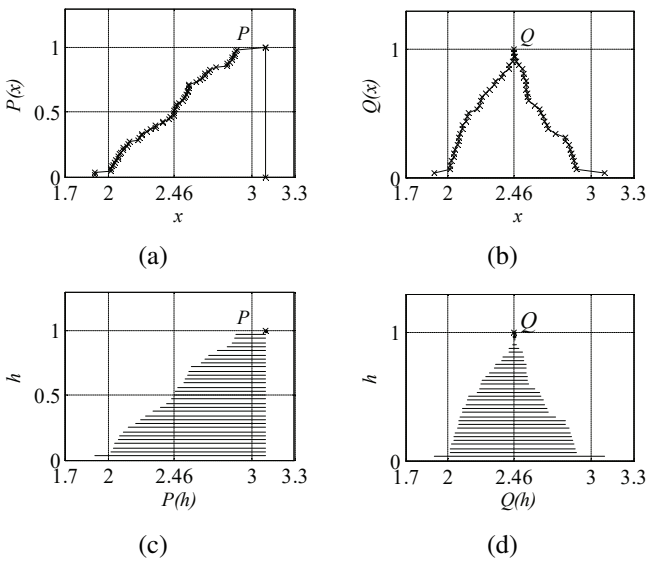

Fig. 1. Intervals' Numbers (INs). (a) A probability distribution IN P "membership-function representation" with horizontal x-axis domain $\mathbb{R}$; 62 samples are indicated by "x". (b) A fuzzy number (possibility distribution) IN Q "membership-function representation" with horizontal x-axis domain $\mathbb{R}$; 62 samples are indicated by "x". (c) The corresponding probability distribution P "interval representation" with $L = 32$ levels and vertical axis domain $[0, 1]$. (d) The corresponding possibility distribution (fuzzy number) Q "interval representation" with $L = 32$ levels and vertical axis domain $[0, 1]$.

where $\sigma_{\sqcap} : \mathbb{I}_1 \times \mathbb{I}_1 \to [0, 1]$ is an *order measure* function given by $\sigma_{\sqcap}([a_1, b_1], [a_2, b_2]; v, \theta) = 1$, if $[a_1, b_1] = \varnothing$; else 0, if $a_1 \vee a_2 > b_1 \wedge b_2$ and $\frac{v(\theta(a_1 \vee a_2)) + v(b_1 \wedge b_2)}{v(\theta(a_1)) + v(b_1)}$, otherwise;

$$\sigma_{\curlyvee}(E, F; v, \theta) = \int_0^1 \sigma_{\sqcup}(E_h, F_h; v, \theta) dh, \tag{6}$$

where $\sigma_{\sqcup} : \mathbb{I}_1 \times \mathbb{I}_1 \to [0, 1]$ is an *order measure* function given by $\sigma_{\sqcup}([a_1, b_1], [a_2, b_2]; v, \theta) = 1$, if $[a_1, b_1] = \varnothing = [a_2, b_2]$ and $\frac{v(\theta(a_2)) + v(b_2)}{v(\theta(a_1 \wedge a_2)) + v(b_1 \vee b_2)}$, otherwise;

$$\rho_1(E, F; v, \theta) = \int_0^1 \varrho_1(E_h, F_h) dh \tag{7}$$

where $\varrho_1 : \mathbb{I}_1 \times \mathbb{I}_1 \to [-1, 1]$ is a *similarity cosine* function given by $\varrho_1([a_1, b_1], [a_2, b_2]) = \frac{a_1 a_2 + b_1 b_2}{\sqrt{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}}$.

Parametric function $v : \mathbb{L} \to \mathbb{R}$ is strictly increasing, whereas parametric function $\theta : \mathbb{L} \to \mathbb{L}$ is strictly decreasing.

D. Cartesian Product Extensions

Let H^N, for an integer N, be the Cartesian product of a pre-Hilbert (inner product) space H with inner product $<,>: H \times H \to \mathbb{R}$. Given $\overline{x} = (x_1, \ldots, x_N)$ and $\overline{y} = (y_1, \ldots, y_N)$ in H^N, let the sum $\overline{x} + \overline{y}$ be defined as $\overline{x} + \overline{y} = (x_1 + y_1, \ldots, x_N + y_N)$; moreover, let the scalar multiple $\kappa \overline{y}$ of $\overline{y} \in H^N$ by $\kappa \in \mathbb{R}$ be defined as $\kappa \overline{y} = (\kappa y_1, \ldots, \kappa y_N)$. Then, the function $<,>: H^N \times H^N \to \mathbb{R}$ defined as

$$< (y_1, \ldots, y_N), (z_1, \ldots, z_N) >= \sum_{i=1}^{N} < y_i, z_i > \quad (8)$$

is an inner product in the *linear (vector) space* H^N. The Hilbert space H could be $H = \mathbb{G}_1$. The following result regards a general lattice including lattice $(\mathbb{G}_1, \preceq)$.

Given an order measure $\sigma_i, i \in \{1, \ldots, N\}$, in a constituent lattice $(\mathbb{L}_i, \sqsubseteq_i)$, an order measure σ_c is defined in the Cartesian product lattice $\mathbb{L} = (\mathbb{L}_1, \sqsubseteq_1) \times \cdots \times (\mathbb{L}_N, \sqsubseteq_N)$ by the following *convex combination* function

$$\sigma_c(\overline{F}, \overline{E}) = k_1\sigma_1(F_1, E_1) + \cdots + k_N\sigma_N(F_N, E_N), \quad (9)$$

where $\overline{F} = (F_1, \ldots, F_N)$ and $\overline{E} = (E_1, \ldots, E_N)$, furthermore $k_1, \ldots, k_N \geq 0$ such that $k_1 + \cdots + k_N = 1$, resulting in a capacity for rigorously fusing disparate types of data.

Given the metric spaces $(X_1, d_1), \ldots, (X_N, d_N)$, it is well known that a family of metric distances can be computed in the Cartesian product $X = X_1 \times \cdots \times X_N$ as

$$d_p(X, Y) = [(d_1(x_1, y_1))^p + \cdots + (d_N(x_N, y_N))^p]^{1/p}, \quad (10)$$

for $N-$tuples $X = (x_1, \ldots, x_N)$ and $Y = (y_1, \ldots, y_N)$.

E. Non-linearities in the space of Intervals' Numbers (INs)

Non-linearities can be introduced in the linear space $\mathbb{G}_1$ and, ultimately, in its convex cone $\mathbb{F}_1$ by parametric functions $v(.)$ and $\theta(.)$ [15].

Given a real function $f : \mathbb{R} \to \mathbb{R}$, let the image (transformation) $f([x_1, \ldots, x_N])$ of a vector of real numbers $[x_1, \ldots, x_N]$ be defined as $[f(x_1), \ldots, f(x_N)]$. If function f is a positive valuation function, say $f = v$, then $[a, b] \in \mathbb{I}_1$ implies $v([a, b]) \in \mathbb{I}_1$. Furthermore, by defining $G_h = v(E_h)$, $h \in [0, 1]$ for an IN $E \in \mathbb{F}_1$, it follows that $G = v(E) \in \mathbb{F}_1$ [15]. The IN $v(E)$ is called the *image* of IN E *filtered* by the positive valuation function v. Furthermore, by defining $v(\mathbb{F}_1)$ as the set $v(\mathbb{F}_1) = \{G : G = v(E), \forall E \in \mathbb{F}_1\}$, it follows that $(v(\mathbb{F}_1), \preceq)$ is a lattice, namely *filtered image* of lattice $(\mathbb{F}_1, \preceq)$. The following results are known [15].

Let $v(.)$ be a positive valuation (i.e., a strictly increasing) function $v : \mathbb{R} \to \mathbb{R}$ such that $v(x) = -v(-x)$, let $\theta(x) = -x$, furthermore let $A, B \in (\mathbb{F}_1, \preceq)$. Then,

(R1) the lattices $(\mathbb{F}_1, \preceq)$ and $(v(\mathbb{F}_1), \preceq)$ are order-isomorphic, symbolically $(\mathbb{F}_1, \preceq) \approx (v(\mathbb{F}_1), \preceq)$, in the sense that 1) $v(.)$ is bijective (i.e. one-to-one and onto), 2) $v(A \curlywedge B) = v(A) \curlywedge v(B)$ and 3) $v(A \curlyvee B) = v(A) \curlyvee v(B)$.

(R2) $D_1(A, B; v, \theta) = D_1(v(A), v(B); x, \theta)$.

(R3) $\sigma_{\curlywedge}(A, B; v, \theta) = \sigma_{\curlywedge}(v(A), v(B); x, \theta)$.

(R4) $\sigma_{\curlyvee}(A, B; v, \theta) = \sigma_{\curlyvee}(v(A), v(B); x, \theta)$.

The results (R2)-(R4) substantiate that lattice $(\mathbb{F}_1, \preceq)$, equipped with both $v(x) = -v(-x)$ and $\theta(x) = -x$, preserves the distances $D_1(.,.)$ as well as the order measures $\sigma_{\curlywedge}(.,.)$ and $\sigma_{\curlyvee}(.,.)$ between bijective elements in lattices $(\mathbb{F}_1, \preceq)$ and $(v(\mathbb{F}_1), \preceq)$. In words, the results (R2)-(R4) guarantee that the shapes of the distributions A and B may be kept intact, thus retaining the corresponding "probabilistic" and/or "possibilistic" interpretations, nevertheless the distance as well as an order measure between the distributions A and B can be tuned non-linearly toward optimizing decision-making.

F. Gradient in the Convex Cone $\mathbb{F}_1$

An attempt to define a derivative in lattice-ordered convex cone $\mathbb{F}_1$ of INs was reported in [28]. The disadvantage of the latter definition was that it might result in GINs outside the set $\mathbb{F}_1$ of INs. This work proposes a theoretically sound approach toward a derivative in $\mathbb{F}_1$ along the following lines. Specifically gradient based updates are defined as:

$$\overline{w}_{k+1} = \overline{w}_k - \eta \triangledown E(\overline{w}_k), \quad (11)$$

where $\overline{w}_{k+1}, \overline{w}_k \in \mathbb{G}_1^N$, $k \in \{0, 1, 2, \ldots\}$ is the (sampling) time, η is a small positive number, $E : \mathbb{G}_1^N \to \mathbb{R}$ is an error function and the gradient (vector) $-\triangledown E(\overline{w}_k) \in \mathbb{G}_1^N$ is the direction of *steepest descent*. In particular, it is $\triangledown E(\overline{w}) = \arg\max_{\|\overline{v}\|=1} DE(\overline{w})[\overline{v}]$, where $DE(\overline{w})[\overline{v}] = \lim_{\varepsilon \to 0} \frac{E(\overline{w}+\varepsilon\overline{v}) - E(\overline{w})}{\varepsilon}$ is the *directional derivative* at $\overline{w}$ in the direction of $\overline{v}$.

By the *Riesz Representation Theorem* [29] for a continuous linear functional $E(\overline{w})$, there exists a unique vector $\triangledown E(\overline{w})$, namely *gradient*, such that $DE(\overline{w})[\overline{v}] =< \triangledown E(\overline{w}), \overline{v} >$.

The abovementioned Riesz Representation Theorem is valid in a Hilbert space e.g. $\mathbb{G}_1^N$, but not necessarily in a convex cone of a Hilbert space e.g. $\mathbb{F}_1^N$. Therefore, $\overline{w}_k \in \mathbb{F}_1^N$ does not guarantee that $\overline{w}_{k+1} \in \mathbb{F}_1^N$. Nevertheless, the gradient provides tools to guarantee that $\overline{w}_{k+1}$ will be in $\mathbb{F}_1^N$ as it will be shown in a future work.

III. Type-1 Neural Network (T1NN)

The proposed neural network architecture is shown in Fig.2. It is a multi-layer feed-forward neural network extension from the Euclidean space $\mathbb{R}^N$ to the space $\mathbb{F}_1^N$ of INs.

The term "Type-1 Neural Networks (T1NNs)" has been proposed for (meta-statistical [30]) deep learning neural networks that process vectors of Type-1 INs in $(\mathbb{F}_1, \preceq)$; whereas, conventional deep learning neural networks that process vectors of numbers in $\mathbb{R}^N$ are called "Type-0 Neural Networks (T0NNs)" [15] – Note that the term "IN Neural Network (INNN)" has been previously used instead of T1NN [30], [31].

A T1NN operates like a T0NN, with two differences: first, a T1NN processes vectors of (Type-1) INs and, second, a link weight may be of two types namely 1) an IN and 2) a strictly monotone function as explained next.

a) Enhanced T1NN Link Weights: The set $\mathbb{R}$ may be interpreted as either a real *vector space* ot a *mathematical field*. On one hand, a T0NN treats $\mathbb{R}$ indiscriminately as both a vector space and a mathematical field. On the other hand, a T1NN separates the aforementioned two interpretations. Hence, two types of link weights result in as it follows. First, a dot product $\overline{w} \cdot \overline{x}$ term $w_i x_i$, $i \in \{1, \ldots, N\}$ in an T0NN may be interpreted as an inner product in $\mathbb{R}$. Second, a dot

product $\overline{w} \cdot \overline{x}$ term $w_i x_i$, $i \in \{1, \ldots, N\}$ may be interpreted as the value of a strictly monotone linear function $f_i(x_i) = w_i x_i$, $i \in \{1, \ldots, N\}$. In both cases, the outcome is the same.

However, for T1NNs the outcomes are different: First, given $\overline{w}, \overline{x} \in \mathbb{F}_1^N$, the dot product $\overline{w} \cdot \overline{x}$ term $w_i \cdot x_i$ is in $\mathbb{R}$. Second, given strictly increasing functions $f_i : \mathbb{R} \to \mathbb{R}$, $i \in \{1, \ldots, N\}$, the result $f_i(x_i)$ is in $\mathbb{F}_1$. The latter interpretation has been enhanced by including strictly decreasing functions $w_i : \mathbb{R} \to \mathbb{R}$, $i \in \{1, \ldots, N\}$ [10], as explained next.

For $w_i < 0$, filtering an IN F, represented by a L-dimentional vector, results in a L-dimentional vector $w_i(F)$ with decreasingly ordered entries. Given a vector $x = [x_1, \ldots, x_N]$, let the entries of vector $r(x) = [y_1, \ldots, y_N] = y$ be defined in reverse order i.e. $y_i = x_{N+1-i}, i \in \{1, \ldots, N\}$. In conclusion, the output vector $r(w_i(F))$ represents an IN [15]. The latter technique is compatible with the multiplication of real numbers in the sense that the result is the same.

A linear weight function $w_i : \mathbb{R} \to \mathbb{R}$, $i \in \{1, \ldots, N\}$ can be replaced by a strictly monotone (either increasing or decreasing) parametric function. Hence, link weight filtering can change not only the location of an IN but also its shape.

Whatever type a link weight v_{ji} might be used in Fig.2, a link's filtered output $v_{ji}(O_{ji})$ is always an (optimizable) IN, where O_{ji} is an output IN of a neuron from the previous layer.

It is noteworthy that the computations in T1NN reduce to computations in T0NN when the input IN $\overline{F} \in \mathbb{F}^N$ is trivial.

b) T1NN Testing: The testing phase of T1NN is straightforward, according to the previous. Moreover, the *sigmoid* function $a : \mathbb{R} \to \mathbb{R}_0^+$ given by

$$\alpha(x; A_\alpha, \lambda_\alpha, \mu_\alpha) = \frac{A_\alpha}{1 + e^{-\lambda_\alpha (x - \mu_\alpha)}} \tag{12}$$

has been used as a neuron's activation function; whereas, the *logistic* function, has been used as a link's weight function:

$$\ell(x; A_\ell, \lambda_\ell, \mu_\ell) = \frac{A_\ell}{2} \frac{1 - e^{-\lambda_\ell (x - \mu_\ell)}}{1 + e^{-\lambda_\ell (x - \mu_\ell)}} \tag{13}$$

A parameter λ_ℓ could be either positive or negative resulting in a strictly increasing or decreasing function, respectively.

An input IN s_j to the activation function of a neuron n_j, $j \in \{1, \ldots, M\}$ is computed as

$$s_j = \left[\sum_{i=1}^{KL} v_{ji}(O_{pji}) \right] + b_j, \tag{14}$$

where $v_{ji}(.)$, $j \in \{1, \ldots, M\}$, $i \in \{1, \ldots, KL\}$ is an incoming link's (from the previous layer of neurons) weight function, O_{pji}, $j \in \{1, \ldots, M\}$, $i \in \{1, \ldots, KL\}$ is the corresponding output IN of a neuron from the previous layer regarding pattern p, and b_j is the bias IN of neuron n_j [15].

c) T1NN Training: The optimization of T1NN's weights is pursued during the training phase. That problem has been solved for T0NN by the (backpropagation) *generalized delta rule* [32] i.e. a steepest descent "hill climbing" algorithm.

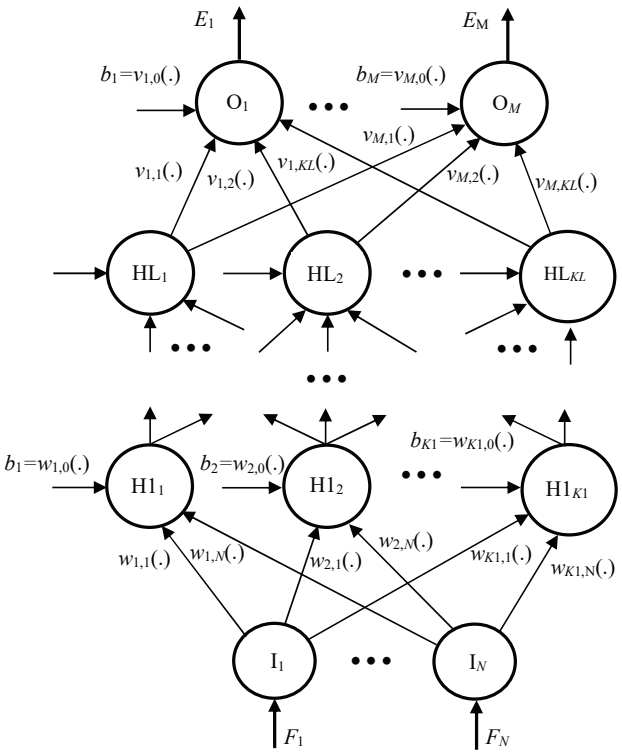

Fig. 2. The $(L+2)$-layer $N \times K1 \times \cdots \times KL \times M$ feed-forward neural architecture in this figure processes INs all along. The weights $v_{i,j}(.)$, where $i \in \{1, \ldots M\}, j \in \{0, \ldots KL\}$, $\ldots$, $w_{i,j}(.)$, where $i \in \{1, \ldots K1\}, j \in \{0, \ldots N\}$ might be either strictly increasing functions or INs. The input to as well as the output from a neuron is an IN. Data processing occurs in the convex cone $\mathbb{F}_1$ of INs in the Hilbert space $\mathbb{G}_1$. Non-linear transformations are carried out by either the link-weights or the transfer functions of neurons.

Previous work with T1NNs has employed rather slow optimization algorithms such as exhaustive search [10] and genetic algorithms [30]. Algorithm 1 below outlines a steepest descent algorithm based on (11) as a combination of backpropagation and the Riesz Representation Theorem.

Algorithm 1 T1NN Training Phase

1: Randomize all T1NN parameters;
2: **while** a termination condition T is not satisfied **do**
3: **for** the next input pattern p **do**
4: Let the gradient $-\nabla E(\overline{w})$ be the δ required to optimize the output layer weights by (11); Backwards calculate the δ required to optimize the weights of all layers successively from the output to the input;
5: **end for**
6: **end while**

The training phase of T1NN computes $\overline{w}_{k+1}$ using (11) under the constraint that $\overline{w}_{k+1} \in \mathbb{F}_1^N$.

When all input link weights in a neuron are interpreted as INs then a fuzzy rule emerges toward local XAI. Axiomatic logic can be engaged in $(\mathbb{F}_1, \preceq)$ by FLR techniques.

IV. Computational Demonstrations

A. IN Induction

IN induction produces T1NN inputs. INs may be induced either from human experts [16], [18], [22], [23] as possibility distributions, i.e. fuzzy sets, or from signals; for the latter, an IN may represent a distribution of measurements e.g. from segmented RGB images [10], [17], [19]–[21] as well as from segmented time series [26]–[28], [30], [31].

B. Computational Experiments

Computational experiments have been carried out to demonstrate the *metric distance* function $D_1 : \mathbb{F}_1 \times \mathbb{F}_1 \to \mathbb{R}_0^+$ of (4), the two *order measure* functions $\sigma : \mathbb{F}_1 \times \mathbb{F}_1 \to [0, 1]$ of (5) and (6), respectively, and the *similarity cosine* function $\rho_1 : \mathbb{F}_1 \times \mathbb{F}_1 \to [-1, 1]$ of (7), as shown in Fig.3.

More specifically, Fig.3(a) displays INs P and Q as well as the assumed positive valuation function $v(x) = \frac{1-e^{-x}}{1+e^{-x}}$; furthermore, it was assumed $\theta(x) = -x$. IN Q was kept constant, whereas IN P was sliding along the horizontal x axis, as indicated by the arrows over P and Q in Fig.3(b).

Fig.3(b) displays the graph of the metric distance $D_1(P, Q)$. In particular, the value of $D_1(P, Q)$ at the location of P and Q in Fig.3(b) is indicated by a "dot" on the vertical line through the left end of the support of IN P (between $x = -6$ and $x = -5$). Fig.3(b) confirms that the distance $D_1(P, Q)$ between P and Q is a convex function with (non-zero) minimum near IN Q. The minimum value of $D_1(P, Q)$ is non-zero because $P \neq Q$ for all locations of the sliding IN P. The values of distance $D_1(P, Q)$ are not symmetrical to the right and to the left of IN Q due to the location of the (saturated) positive valuation function $v(x) = \frac{1-e^{-x}}{1+e^{-x}}$.

Fig.3(c) displays the graphs of the order measure functions $\sigma_\curlywedge(P, Q)$ and $\sigma_\curlyvee(P, Q)$, respectively. Note that, as expected from their definitions, Fig.3(c) confirms that $\sigma_\curlyvee(P, Q) \neq 0$, $\forall x$, whereas $\sigma_\curlywedge(P, Q) = 0$ for non-overlapping INs P and Q.

Fig.3(d) displays the graph of the similarity cosine $\rho_1(P, Q)$. It is noteworthy that there is an x_0 such that $\rho_1(P, Q) = 0$ i.e. $< P, Q >= 0$ meaning that the INs P and Q are orthogonal to one another at x_0; hence, established tools become available in $\mathbb{F}_1^N$ for rigorous mathematical analysis.

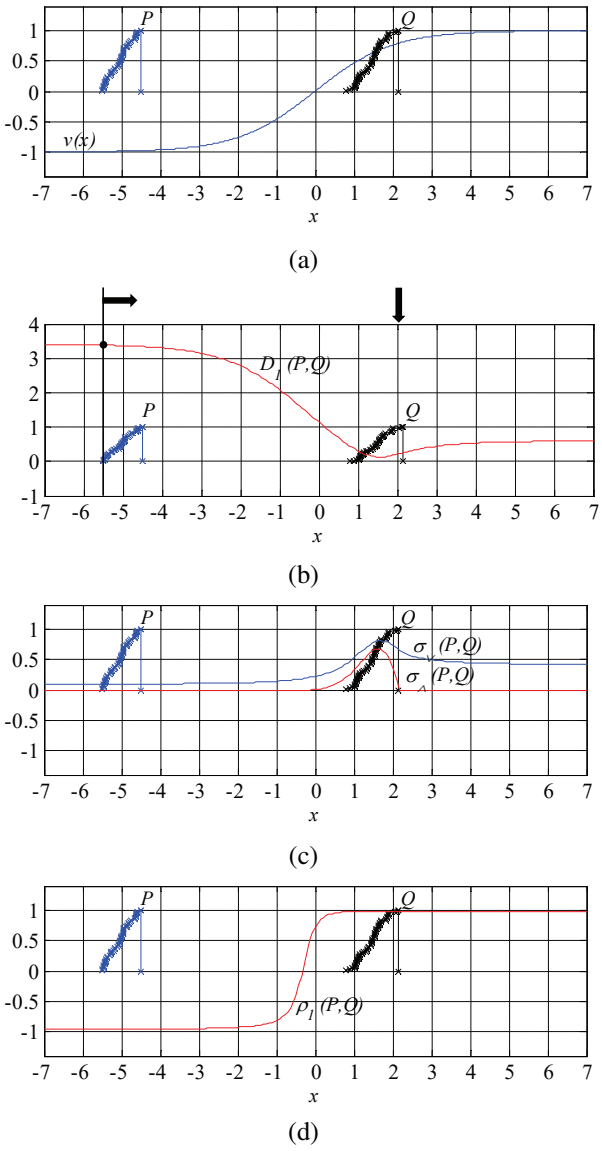

Fig. 3. (a) INs P and Q and the positive valuation function $v(x) = \frac{1-e^{-x}}{1+e^{-x}}$; it is assumed $\theta(x) = -x$. (b) Graph of the metric distance function $D_1(P, Q)$ between the sliding IN P and the still IN Q. The "dot" on the graph indicates the value of $D_1(P, Q)$ at the current position of the sliding IN P. (c) The blue and red graphs show the order measure functions $\sigma_\curlyvee(P, Q)$ and $\sigma_\curlywedge(P, Q)$, respectively. (d) Graph of the similarity cosine function $\rho_1(P, Q)$. For $\rho_1(P, Q) = 0 =< P, Q >$, INs P and Q are orthogonal to one another.

V. Discussion and Conclusion

This section discusses, first, the motivation for the proposed mathematical instruments, second, the contribution of this work and, third, potential future work extensions.

A. Mathematical Background Motivation

The notion "Hilbert space" has been introduced by David Hilbert [33], also known for his influential work on the foundations of geometry [34]. Nevertheless, the formal definition of a Hilbert space was proposed by von Neumann, who applied it instrumentally in quantum mechanics [35].

Two alternative mathematical structures for computing include, first, a Riesz space, i.e. a lattice-ordered vector space [29], and, second, a Banach space, i.e. a complete normed vector space [36]. An "ontological" difference between a Hilbert space and a Riesz / Banach space, is the existence of an inner product in a Hilbert space. Even though, compared to a Riesz / Banach space, a Hilbert lattice is defined more restrictively, its close mathematical relation to the Euclidean space [11] makes a Hilbert space a promising candidate for enhanced practical applications regarding deep learning AI.

Reproducing Kernel Hilbert Space (RKHS) is a machine learning approach that depends on a Hilbert space of functions without involving lattice theory [37] as the present work is doing. The objective of this work is to enhance incrementally the commercially successful deep learning technology.

B. The Contribution

The techniques proposed here fall within the *Lattice Computing* (LC) *information processing paradigm*, which has been defined as "an evolving collection of tools and methodologies that process lattice ordered data including logic values, numbers, sets, symbols, graphs, etc" [9], [38].

Previous work has set the goal of defining a derivative function in the Hilbert space $\mathbb{G}_1$ toward the development of fast-converging T1NN training algorithms. As the notion of gradient is inherent in a Hilbert space, Algorithm 1 points out to a promising direction for T1NN training.

This position paper has proposed two necessary and sufficient conditions for neural computing including, first, a Hilbert space H and, second, non-linear transformations in H according to the following rationale.

First, a Hilbert space H, by definition is 1) a *linear space*, 2) equipped with an *inner product* and 3) *complete* in the sense that every Cauchy sequence in H converges in H. It is well known that the popular Euclidean space $\mathbb{R}^N$ is a Hilbert space; moreover, the operations of the Hilbert space $\mathbb{R}^N$ and only those operations are used in conventional (T0NN) neural computing. Specifically, the "dot product" input to a neuron is the inner product in Hilbert space $\mathbb{R}^N$; the "similarity cosine", used extensively in deep learning, is defined in a Hilbert space; furthermore, the "completeness" of a Hilbert space is a criterion for training convergence in deep learning. Therefore, neural computing was extended from T0NN in Hilbert space $\mathbb{R}^N$ to T1NN in the convex cone $\mathbb{F}_1^N$ of the Hilbert space $\mathbb{G}_1^N$.

Second, non-linear transformations are typically introduced in T0NN deep learning by a neuron's transfer function. In addition, non-linearities can be introduced in T1NN deep learning by the link weights, be it INs or strictly monotone functions, as well as by the underlying functions $v(.)$ and $\theta(.)$.

The Hilbert space $\mathbb{G}_1$, of GINs, was presented as a hierarchy of Hilbert spaces stemming from $\mathbb{R}$. Then, the interest focused on the convex cone $\mathbb{F}_1 \subset \mathbb{G}_1$ of INs, where an IN represents an (information) granule. Specifically, INs represent possibility /probability distributions including real numbers. Furthermore, $(\mathbb{F}_1, \preceq)$ is a sublattice of lattice $(\mathbb{G}_1, \preceq)$.

The set $\mathbb{F}_1$ of INs is a domain for potential unification and/or enhancement of conventional neural- and/or fuzzy- systems. Note that the architecture in Fig.2 can be interpreted as a fuzzy inference system (FIS) [18]. In such a context, the similarity cosine $\rho_1(.,.)$ in (7) suggests a novel instrument for calculating the activation of fuzzy rules.

Given that $\mathbb{F}_1 \supset \mathbb{R}$, it follows that T1NNs can achieve as much as T0NNs can achieve and, more than that, they can enhance T0NNs by processing information granules. Hence, a far-reaching potential of T1NNs emerges, that is computing with semantics, represented by partial-order, instead of merely "number crunching". Axiomatic logic can be engaged in $(\mathbb{F}_1, \preceq)$ by FLR toward XAI. Enhanced embeddings of symbols emerge based on the enhanced T1NN link weights.

There are set-theoretic bounds on deep learning due to the uncountably infinite cardinality $\aleph_1$ of the set $\mathbb{F}_1$ of INs [22]. In particular, first, the cardinality of any parametric family of models, including T0NNs, equals $\aleph_1$ [23]; second, the cardinality of the set of functions $f : \mathbb{F}^N \rightarrow \mathbb{F}^M$ with a capacity for generalization, including XAI rule-based surrogate T1NNs [39], equals $\aleph_2 = 2^{\aleph_1} > \aleph_1$ [22], [23]. However, due to the finite "digital word length", only a finite number of models can be implemented on a digital computer.

As $\mathbb{F}_1 \supset \mathbb{R}$, the representation potential of T1NN is superior to that of T0NN. In particular, note that, even though the sets $\mathbb{F}_1$ and $\mathbb{R}$ have equal cardinalities, symbolically $|\mathbb{F}_1| = \aleph_1 = |\mathbb{R}|$, an element of $\mathbb{R}$ represents a single real number, whereas an element of $\mathbb{F}_1$ may represent up to countably infinite $\aleph_0$ real numbers (i.e., *all order* data statistics [16]) using only $L < +\infty$ real numbers; the latter is an advantage of using INs in big data applications resulting in, potentially, smaller computer memory requirements. Furthermore, no *ad hoc* feature extraction is necessary [21]. It is known that T0NN deep learning training typically forces "nearby" objects of interest to be represented by "nearby" N-dimensional vectors of real numbers [40]. As the set $\mathbb{F}_1$ of INs is equipped with the same mathematical instruments it is expected that T1NN deep leaning training can also force "nearby" objects of interest to be represented by "nearby" N-dimensional vectors of INs.

C. Potential Future Work

Several theoretical and practical questions remain to be addressed including the development of fast-convergence training algorithms for the T1NNs according to Algorithm 1. An alternative introduction of non-linearities may be effected by defining a different inner-product at each different element of the Hilbert space $\mathbb{G}_1$, thus defining a Riemannian manifold while keeping intact the shapes of the INs involved.

An ever better performance is reported in the literature for models with an ever larger number of parameters [13]. T0NNs use no more than one parameter (weight) per link between neurons, whereas a T1NN can have any number of parameters per link between two neurons. Hence, a T1NN with fewer neurons can have the same number of parameters as a T0NN with more neurons. Note that previous works have demonstrated that the employment of T1NN/INNN has comparatively demonstrated a clearly better performance in a number of applications [17], [20], [21], [28], [30], [31]. Statistical- and/or other testing needs to confirm any improvement of performance by T1NNs. In addition, as a T1NN can have the same number of parameters as a T0NN using fewer neurons, the potential of T1NNs to retain performance with less electric energy consumption also needs to be tested. Another topic for future work is the development of T2NNs for processing Type-2 INs including Type-2 fuzzy sets [41].

Hilbert spaces are indispensably used in quantum computing [42]. Based on instruments presented in this paper, future work will consider quantum implementations of neural models.

The use of symbols is instrumental in a spoken language e.g. for reasoning/planning. Potential future work also includes the development of a series of successive (lattice-order) isomorphisms from signals to symbols toward grounding symbols [43] such that semantics is successively retained.

REFERENCES

[1] I. Newton, *The Mathematical Principles of Natural Philosophy (Philosophiae Naturalis Principia Mathematica)*. Createspace Independent Publishing Platform, 2018.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, published 28 May 2015. [Online]. Available: https://www.nature.com/articles/nature14539

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and L. Kaiser. (2017) Attention is all you need. [Online]. Available: https://arxiv.org/abs/1706.03762

[5] G. Boole, "The calculus of logic," *Cambridge and Dublin Mathematical Journal*, vol. III, pp. 183–198, 1848.

[6] L. A. Zadeh, "Fuzzy sets," *Info. and Control*, vol. 8, pp. 338–353, 1965.

[7] G. Birkhoff, *Lattice Theory*, ser. Colloquium Publications. Providence, RI: American Mathematical Society, 1967, vol. 25.

[8] G. X. Ritter and G. Urcid, *Introduction to Lattice Algebra with Applications in AI, Pattern Recognition, Image Analysis, and Biomimetic Neural Networks*. Boca Raton, FL: Chapman and Hall/CRC, 2021.

[9] V. G. Kaburlasos, "Lattice computing: A mathematical modelling paradigm for cyber-physical system applications," *Mathematics*, vol. 10, no. 2, p. 271, 16 January 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/2/271

[10] V. Kaburlasos and G. Siavalas, "Computing with semantics in a Hilbert space: a proposal for enhancing deep-learning," in *Biologically Inspired Cognitive Architectures (BICA) 2025*, ser. Studies in Computational Intelligence, A. V. Samsonovich, F. Ramos, and T. Liu, Eds. Cham, CH: Springer, 2026, vol. 1243, pp. 311–325.

[11] W. Rudin, *Functional Analysis*, 2nd ed., ser. International Series in Pure and Applied Mathematics. Hamburg, Germany: McGraw-Hill, 1991.

[12] V. G. Kaburlasos, *Towards a Unified Modeling and Knowledge-Representation Based on Lattice Theory*, ser. Studies in Computational Intelligence. Heidelberg, Germany: Springer, 2006, vol. 27.

[13] V. G. Kaburlasos, C. Lytridis, E. Vrochidou, C. Bazinas, G. A. Papakostas, A. Lekova, O. Bouattane, M. Youssfi, and T. Hashimoto, "Granule-based-classifier (GbC): a lattice computing scheme applied on tree data structures," *Mathematics*, vol. 9, no. 22, p. 2889, 13 November 2021. [Online]. Available: https://www.mdpi.com/2227-7390/9/22/2889

[14] Reuters. https://www.reuters.com/business/apple-use-googles-ai-model-run-new-siri-bloomberg-news-reports-2025-11-05/.

[15] V. Kaburlasos, G. Siavalas, and L. Bris, "A hierarchy of Hilbert spaces for computing with granular semantics by deep learning enhanced," in *Computational Semantics - Bridging Language, Logic, and Learning*, J. Terven and D. M. C. Esparza, Eds. London, U.K.: IntechOpen Ltd, published 07 January 2026, ch. 1, pp. 1–27. [Online]. Available: https://www.intechopen.com/books/1004462

[16] S. E. Papadakis and V. G. Kaburlasos, "Piecewise-linear approximation of nonlinear models based on probabilistically/possibilistically interpreted Intervals' Numbers (INs)," *Information Sciences*, vol. 180, no. 24, pp. 5060–5076, 2010.

[17] V. G. Kaburlasos, S. E. Papadakis, and G. A. Papakostas, "Lattice computing extension of the FAM neural classifier for human facial expression recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1526–1538, 2013.

[18] V. G. Kaburlasos and A. Kehagias, "Fuzzy inference system (FIS) extensions based on the lattice theory," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 3, pp. 531–546, 2014.

[19] V. G. Kaburlasos and G. A. Papakostas, "Learning distributions of image features by interactive fuzzy lattice reasoning (FLR) in pattern recognition applications," *IEEE Computational Intelligence Magazine*, vol. 10, no. 3, pp. 42–51, 2015.

[20] V. G. Kaburlasos, G. A. Papakostas, T. Pachidis, and A. Athinellis, "Intervals' numbers (INs) interpolation /extrapolation," in *Proc. IEEE Intl. Conf. on Fuzzy Systems*, Hyderabad, India, 7-10 July 2013.

[21] V. G. Kaburlasos, E. Vrochidou, C. Lytridis, G. A. Papakostas, T. Pachidis, M. Manios, S. Mamalis, T. Merou, S. Koundouras, S. Theocharis, G. Siavalas, C. Sgouros, and P. Kyriakidis, "Toward big data manipulation for grape harvest time prediction by intervals' numbers techniques," in *Proc. World Congress on Computational Intelligence (WCCI) 2020, Intl. J. Conf. on Neural Networks (IJCNN 2020) Program*, Glasgow, Scotland, U.K., 19-24 July 2020.

[22] V. G. Kaburlasos and A. Kehagias, "Novel fuzzy inference system (FIS) analysis and design based on lattice theory. Part I: working principles," *Intl. J. of General Systems*, vol. 35, no. 1, pp. 45–67, 2006.

[23] ——, "Novel fuzzy inference system (FIS) analysis and design based on lattice theory," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 2, pp. 243–260, 2007.

[24] A. L. Ralescu and D. A. Ralescu, "Probability and fuzziness," *Information Sciences*, vol. 34, no. 2, pp. 85–92, 1984.

[25] S. Wonneberger, "Generalization of an invertible mapping between probability and possibility," *Fuzzy Sets and Systems*, vol. 64, no. 2, pp. 229–240, 1994.

[26] V. G. Kaburlasos, "FINs: lattice theoretic tools for improving prediction of sugar production from populations of measurements," *IEEE Trans. Systems, Man and Cybernetics – B*, vol. 34, no. 2, pp. 1017–1030, 2004.

[27] V. Kaburlasos, E. Vrochidou, F. Panagiotopoulos, C. Aitsidis, and A. Jaki, "Time series classification in cyber-physical system applications by intervals' numbers techniques," in *Proc. FUZZ-IEEE 2019*, New Orleans, LA, 23-26 June 2019.

[28] V. G. Kaburlasos, C. Bazinas, E. Vrochidou, and E. Karapatzak, "Agricultural yield prediction by difference equations on data-induced cumulative possibility distributions," in *Applications of Fuzzy Techniques*, ser. Lecture Notes in Networks and Systems (LNNS), S. Dick, V. Kreinovich, and P. Lingras, Eds. Cham, CH: Springer, 2022, vol. 500, pp. 90–100.

[29] W. A. J. Luxemburg and A. C. Zaanen, *Riesz Spaces*. Amsterdam, The Netherlands: North-Holland, 1971.

[30] C. Bazinas, C. Lytridis, and V. G. Kaburlasos, "Meta-statistical deep learning for stochastic time-series prediction in agricultural applications," in *Proc. 2023 Intl. Neural Network Society (INNS)*, C. Jayne, D. P. Mandic, and R. J. Duro, Eds. Elsevier, 2023, vol. Procedia Computer Science 222, pp. 293–302. [Online]. Available: https://dblp.org/db/conf/inns-dlia/inns-dlia2023.html

[31] C. Bazinas, E. Vrochidou, C. Lytridis, and V. G. Kaburlasos, "Time-series of distributions forecasting in agricultural applications: an intervals numbers approach," in *Engineering Proceedings*, I. Rojas, F. Rojas, L. J. Herrera, and H. Pomares, Eds., 2021, vol. 5, no. 1, p. 12. [Online]. Available: https://www.mdpi.com/2673-4591/5/1/12

[32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing – Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, Eds. Cambridge, MA: The MIT Press, 1986, vol. 1: Foundations, ch. 8, pp. 318–362.

[33] D. Hilbert, *Grundzge einer allgemeinen Theorie der linearen Integralgleichungen (Foundations of the General Theory of Linear Integral Equations)*. Leipzig, Germany: B. G. Teubner, 1912.

[34] ——, *Grundlagen der Geometrie (Foundations of Geometry)*. Leipzig, Germany: B. G. Teubner, 1899.

[35] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press, 1955.

[36] E. Esmi, L. C. de Barros, F. S. Pedro, and B. Laiate, "Banach spaces generated by strongly linearly independent fuzzy numbers," *Fuzzy Sets and Systems*, vol. 417, pp. 110–129, 2021.

[37] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, UK: Cambridge University Press, 2004.

[38] P. Sussner and I. Campiotti, "Extreme learning machine for a new hybrid morphological/linear perceptron," *Neural Networks*, vol. 123, pp. 288–298, 2020.

[39] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, and K. S. Kwak, "A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimers disease," *Scientific Reports*, vol. 11, p. 2660, 29 January 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-82098-3

[40] A. Pak, A. Ziyaden, T. Saparov, I. Akhmetov, and A. Gelbukh, "Word embeddings: A comprehensive survey," *Computacion y Sistemas*, vol. 28, no. 4, pp. 2005–2029, 2024.

[41] P. Sussner, "Complete lattices as frameworks for interval and general type-2 fuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 9, pp. 2972–2986, 2025.

[42] M. Keçeci, "Understanding quantum mechanics through Hilbert spaces: Applications in quantum computing," *Open Science Articles (OSAs)*, vol. 1, no. 1, pp. 1–41, published 15 August 2025. [Online]. Available: https://doi.org/10.5281/zenodo.15468754

[43] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena*, vol. 42, no. 1-3, pp. 335–345, 1990.