A Temporal Type-2 Fuzzy System for Time-Dependent Explainable Artificial Intelligence

Mehrin Kiani, Member, IEEE, Javier Andreu-Perez, Senior Member, IEEE, and Hani Hagras, Fellow, IEEE

Abstract—Explainable artificial intelligence (XAI) focuses on transparent AI models and decisions, which are easy to understand, analyze, and augment by a nontechnical audience. Fuzzy logic systems (FLS)-based XAI provides an explainable framework while also modeling uncertainties in real-world environments. However, most real-life processes are not characterized by high uncertainty alone; they are also inherently time dependent, i.e., the processes are time variant. In this work, we present a novel temporal type-2 FLS-based approach for time-dependent XAI (TXAI) systems, which can account for the likelihood of a sample occurrence in the time domain by its the frequency. In the proposed temporal type-2 fuzzy sets (TT2FSs), a 4-D time-dependent membership function integrates the universe of discourse, its membership, and its frequency of occurrence across time. The TXAI system manifested better classification prowess in cross-validation tests, with a mean recall of 95.40% than a standard XAI system (based on nontemporal general type-2 fuzzy sets) that had a mean recall of 87.04%. TXAI also performed significantly better than most nonexplainable AI systems, with between 3.95% and 19.04% improvement gain in mean recall. In addition, TXAI can also outline the most likely time-dependent trajectories using the frequency and time dimensions embedded in the TXAI model; viz. given a rule at a determined time interval, what will be the next most likely rule at a subsequent time interval. In this regard, the proposed TXAI system can have profound implications for delineating the evolution of real-life time-dependent processes, such as behavioral or biological processes.

Impact Statement—In this work, the authors introduce a Time-dependent eXplainable Artificial Intelligent (TXAI) system, based on novel Temporal type-2 fuzzy sets (TT2FSs) to analyse real-life processes accounting for their time-dependencies. The membership function (MF) associated with TT2FS, called Temporal MF (TMF), can integrate the time based variation in a fuzzy concept, as well as the variation in the feature domain, integrating a fuzzy relation between time and the concept domain. To the best of the authors’ knowledge, no prior work has undertaken the integration of time-based variation in a fuzzy concept for the computation of its associated MF. Type-2 fuzzy sets and systems in their standard form renders them ill-suited for the time-based analysis of real-life processes. The temporal information integrated in TMF, associated with TT2FS which in-turn constitute the TXAI system, are able to analyse temporal dynamics of a real-life process whilst retaining the hallmark explainability characteristics of a fuzzy logic based XAI system. In addition to the allowance of a time-based analysis, the proposed TXAI systems are also able to delineate the evolution of a real-life process using temporal trajectories. The prowess of TXAI system to shed light on the evolution of a process using temporal trajectories is of paramount significance because it (trajectories) can inform us about the dynamics of a process across time. Delineating the prototypical trajectories of patterns across time domain can have profound implications for in real-word context affected by human dynamics (e.g. ambient intelligence, social systems), or biological processes (e.g. brain function, systems biology).

Index Terms—Artificial intelligence, explainable artificial intelligence, fuzzy systems, human computer interaction, human in the loop, time-varying, trusted computing.

I. INTRODUCTION

Over the last few decades, the widespread application of artificial intelligence (AI) systems have enhanced many aspects of everyday life from risk management [1], sky herding of sheep [2], medical image segmentation [3], recognition of expertise level [4], mobile applications [5] to Covid-19 detection based on cough samples [6]. Although opaque AI systems offer remarkable prediction accuracy, they are limited by a lack of explanation behind their predictions. A lack of explanation renders the AI systems untrustworthy, and particularly inapplicable where users want to understand the decision process of the AI system. To this end, there is a growing need for transparent, human-understandable AI systems called explainable AI (XAI) systems [7]. Several approaches taken toward the development of XAI systems include the following: 1) Intrinsic: A method in which model inference structure is fully transparent, such as short decision trees (DTs) or sparse linear models, and 2) Post-hoc: A model-agnostic meta-model is used to decipher the inference rationale of a black-box model permutation feature importance can be computed for DTs. Within post-hoc methods attempts to unravel a black-box model into a surrogate intrinsic model have also been undertaken. A particular category of these are the anchor-based models.

Although anchor-based approach provides a step toward implementing human-understandable explanations [8], explanatory patterns rest on hard thresholds and are constrained by Boolean logic. However, real-life processes are characterized with uncertainty and, therefore, hard thresholds-based models are not particular well-suited to model them (real-life processes). In this regard, another approach to implement XAI systems is fuzzy logic systems (FLS) [7], [9]. The FLS-based XAI systems are well-suited for explainable modeling of real-life processes.
because of FLS capability to handle uncertainty in the input data, and subsequently improve the process model and performance. In addition, the use of conceptual labels (CoLs) that model uncertainty and axioms of FLS-based XAI systems pave way for human-understandable models for describing complex, real-life processes.

The FLS-based XAI systems handle uncertainty in the input data using fuzzy sets that convert crisp numbers (viz. uncertain observations) to CoLs characterized with membership values [9], [10]. The fuzzy sets are defined by membership functions (MFs) and represent a given CoL. The membership value is usually in the range [0,1] and is a soft measure of the degree of association the associated fuzzy set has for a given crisp measurement to belong to the CoL represented by the fuzzy set [10]. For example, an XAI system modeling the heights of people in a community using type-1 (T1) fuzzy sets may represent height using CoLs of tall, medium, and short. The MF associated with each CoL’s MF will assign a crisp number for the height of a person with a membership grade; for example, a height of 6 ft may get assigned membership grades of 0.8, 0.5, 0.1 to represent CoLs of tall, medium, and short, respectively.

In general, fuzzy sets can model uncertainty in the feature domain at different levels: T1, interval type-2 (IT2), and general type-2 (GT2) fuzzy sets; illustrated in Fig. 1. Despite the variability in the extent for uncertainty modeling amongst the types of fuzzy sets, all fuzzy sets are modeling uncertainty from a single time snapshot of the feature domain. More specifically, fuzzy sets do not integrate associated temporal information in their membership grade calculation. This is a critical limitation of the fuzzy sets since most real-life systems are time-variant, i.e., their behavior changes with time. To model time-dependent real-life systems more effectively, in this work, we present the theory of a new temporal type-2 fuzzy set (TT2FS) based approach for time-dependent XAI (TXAI).

The prowess of TXAI system for incorporating time information for modeling time-variant processes is of paramount significance since the insights provided by a TXAI system can shed light on both spatial (feature domain) and temporal behavior of the time-dependent process. More specifically, the TXAI is able to inform not only about the relation between input features but can also describe the impact of time on the evolution of the inter-relation of the features. As an example, let us consider a standard XAI system composed of a T1 fuzzy set for modeling thermal sensation “cold” in the domain of values of temperature T °C, as shown in Fig. 2(a), and a T1 fuzzy set for the time of occurrence of concept “cold” during the months of a year, as shown in Fig. 2(b). The notion is that the perception of “cold” is mostly associated with the months of winter than in the months of spring. Hence, using the time information associated with a fuzzy concept (such as cold in this case), a temperature can belong to the concept (cold) differently according to a particular point in time (e.g., months of a year).

Crediting a fuzzy membership with its associated time information is particularly advantageous for the modeling of time-dependent noise-prone processes. Moreover, for dynamic processes, the ability to delineate its’ (dynamic process) trajectories across time would inform the evolution of the temporal dynamics of the process. To this end, our proposed TXAI system has been designed to integrate temporal information as well as able to outline the trajectories of a time-dependent process. To demonstrate the efficacy of TXAI system for time-dependent process modeling, in this work, an occupancy dataset is used [11]. Using the values of temperature, light, and carbon dioxide (CO₂), and the time the aforementioned measurements are taken, the TXAI system is used to make a prediction of whether or not the room is occupied.

The rest of this article is organized as follows. In Section II, related works are outlined, Section III presents the TXAI system definition and operations, Section IV outlines the TXAI inference system (TXAI-IS) with a numerical step-by-step example as well as the evolution of a TXAI model using temporal trajectories. An empirical study using TXAI system, as well as state-of-the-art systems (with varying levels of explainability)
for the performance comparison, on the aforementioned occupancy dataset [11] is presented in Section V. Finally, Section VI concludes this article.

II. RELATED WORKS

Fuzzy sets have enabled explainable models of complex real-life processes, which prove too ill-defined for closed-form mathematical analysis. In this regard, although uncertainty in complex processes could be handled by fuzzy sets, the time-variant characteristics of complex processes have not been integrated into the modeling by standard XAI systems based on state-of-the-art fuzzy sets.

There have been few notable attempts in the literature to model time in the MFs. The work by Garibaldi et al. [12] on nonstationary fuzzy sets proposed that variation within an MF can be incorporated by perturbing the parameters of the MF. Their work aims to develop nondeterministic fuzzy reason as a way to model the variability in fuzzy decision making to mimic the variability in expert opinions. The ability of nonstationary fuzzy sets to integrate differing experts’ opinions is a significant contribution since it allows for a more comprehensive model that takes into account all experts’ opinions. However, their work does not incorporate the variation within a fuzzy concept with respect to time, which is the aim of this work, to represent the time-variant transformation of a same fuzzy linguistic variable.

Similarly, the work by Kostikova et al. [13] proposed dynamic fuzzy sets by extending the classical fuzzy set to include a time dimension for representing MF at different time points. They propose four different types of dynamic MFs depending on how many parameters are changed in the definition of the dynamic MF. They simulated their dynamic MFs by using differing expert assessments on multilevel fuzzy description of a complex system. However, the dynamic MF is essentially a set of functions determined at different time points with no bearing on the temporal variation in the fuzzy concept.

In another work by Maeda et al. [14], they propose dynamic fuzzy reason to deal with the notion of time delay between premise and consequent. An example of where a time delay between premise and consequent assumes critical importance is: “If it starts snowing, the traffic on road will increase about 30 min later.” They propose the use of fuzzy relations between a fuzzy concept and its fuzzy time interval to assign a credit degree to the concept. The temporal fuzzy reasoning provides a framework for modeling delay in fuzzy reasoning and the temporal dynamics of a fuzzy concept. In this work, we have built on the work of Maeda et al. [14] to credit the membership grade of a concept based on time.

To the best of the authors’ knowledge, there is no work in the literature on fuzzy sets that delineates the incorporation of time-based variation in a fuzzy concept to compute the membership grade for the crisp values of the fuzzy concept. In addition, no previous work has aimed at delineating the trajectories of a time-variant process with respect to time. To this end, in this work, we propose TXAI systems that can integrate information from both the feature domain and time domain. More details on the proposed TXAI are outlined in Section III.

III. TXAI SYSTEMS

In this section, we present the TXAI system based on TT2FSs that incorporate information from not only the uncertainty in the input domain of the fuzzy linguistic term, but also from its time of occurrence. In particular, the information from the time of occurrence is integrated into the membership grade of the TT2FS using fuzzy relations such that it (the membership grade of the TT2FS) varies with respect to time (time-dependent).

In the following section, we present the most common fuzzy relations and outline how they can be used for implementing TT2FS.

A. Fuzzy Relations Between Fuzzy Linguistic Variables and Time Related Measures

In this work, fuzzy relations are used to interrelate the information with respect to the degree of truth of a determined linguistic term or CoL, A, within the domain X, and time T to form TT2Fs such that the likelihood of occurrence of A in \( x \in X \), i.e., the primary membership grade \( \mu_A(x) \), is credited by a measure that is dependent on time, such as frequency. The application of fuzzy relation, for constructing TT2Fs,
Before reviewing the different relations that can be applied to construct a TT2FS, the conditions that need to be fulfilled by the associated TMF are listed as follows.

i) The TMF should be continuous.

ii) The TMF should be convex.

iii) The range of the TMF ⊆ [0, 1].

iv) The TMF should reflect in the value of membership grade the intrinsic magnitudes of membership grade in feature domain and in frequency of occurrence domain, i.e., they should be directly proportional. For example, if μ_A(x) is high and the time representation is also high then the result from the relation between them should also be high and vice versa.

An illustrative comparison of the TT2FSs formed for the CoL “cold” of feature thermal concept using the fuzzy relations listed in Table I is shown in Fig. 3. The fuzzy relations are applied on hypothetical primary MF of “cold” in feature domain (temperature) and time domain (months of a year). As can be seen in Fig. 3, the different fuzzy relations are encapsulating distinct interdependencies between time and feature domain. All relations meet the criteria i)–iii) listed above, however, only the Mamdani relation meets the criterion iv) as well since it gives credit to μ_Cold based on the variable frequency of occurrence of “cold” as observed in different months of the year. Hence, in this work, the Mamdani relation is used to construct the TT2FSs.

B. Conditional Relative Frequency Distribution of a Fuzzy Linguistic Term

In our TT2FS, we employ a measure of conditional relative frequency between time and the occurrence of a linguistic term. We denote as A an instance of a linguistic term from a set of CoLs (also called words of the universe of discourse), CoLs := [CoL_1, CoL_2, ..., CoL_J] of a specific linguistic variable or input.

**Definition III.1 (Discrete conditional relative frequency with respect to time):** The discretized conditional relative frequency is defined as the likelihood of observing a linguistic term A based on its membership grade, across time. This is denoted as g_A(t_n, μ_A(x)) with time t discretized over N time points (t_n) such as t_n ∈ [t_1, ..., t_N], and is given by

\[ g_A(t_n, μ_A(x)) = \frac{\sum_{x \in X,t_n} δ_{n,j}}{\max_{[t_1, ..., t_N]} \left( \sum_{x \in X,t_n} δ_{n,j} \right)} \]

where δ_{n,j} is a Kronecker delta function [15] (e.g., δ_{ab} = 0 if a ≠ b, δ_{ab} = 1 if a=b) that takes the value of 1 when the following condition applies, \( \exists \ \arg\max \{μ_{Col_j}(x^{n+})\} : Col_j = A, \forall j \in [1, ..., J], \) and 0 otherwise. Note x^{n+} is a realisation of x at time t_n.

The numerator in (1) finds the count of occurrences of a given A for a determined time point t_n across all data instances, whereas the denominator is finding the maximum value of the count of occurrences of A across all N time points and all data instances. The resultant discrete conditional relative frequency \( g_A(t_n, μ_A(x)) \) is interpolated to form a conditional distribution \( f_A(t, μ_A(x)) \). For the sake of notational simplicity, we denote the later distribution as \( f_A \) and the discrete conditional relative frequency as \( g_A \) from here onwards.

**TABLE I**

| Name | Definition of the relation |
|------|-----------------------------|
| Godel | \( R_G(t,x) = \begin{cases} 1, & \text{if } \mu_T(t) \leq \mu_A(x) \\ \mu_A(x), & \text{if } \mu_T(t) > \mu_A(x) \end{cases} \) |
| Lukasiewicz | \( R_L(t,x) = 1 \times (1 - \mu_T(t) + \mu_A(x)) \) |
| Gaines–Rescher | \( R_{GR}(t,x) = \begin{cases} 1, & \text{if } \mu_T(t) \leq \mu_A(x) \\ 0, & \text{if } \mu_T(t) > \mu_A(x) \end{cases} \) |
| Mamdani | \( R_M(t,x) = \mu_T(t) \times \mu_A(x) \) |

Fig. 3. Comparison of TT2FSs for the CoL “cold” for feature thermal concept constructed with the most commonly used fuzzy relations namely Mamdani, Zadeh/Lukasiewicz, Godel, and Gaines–Rescher, see Table I for their respective definitions. In these illustrative plots, the feature domain, i.e., temperature in °C is plotted on the x-axis, with conditional distribution, \( f_{Cold} \) on y-axis, and the time is plotted on the axis connecting the x- and y-axis, i.e., the arc axis, with the 4 time intervals representing the typical seasons in a year. The z-axis has the values of TMF, \( μ_{Cold}(x,t,f_{Cold}) \).
Let us assume that the linguistic variable is thermal sensation defined on the input domain \((x \in X)\) of temperature in °C and the associated CoLs be: [cold, comfortable, hot]. For a given crisp input of temperature such as 15 °C, the associated primary membership grade for all three CoLs of cold, comfortable, and hot be \(\mu_{\text{cold}}(15 \, ^\circ\text{C}) = 0.4\), \(\mu_{\text{comfortable}}(15 \, ^\circ\text{C}) = 0.3\), \(\mu_{\text{hot}}(15 \, ^\circ\text{C}) = 0\), respectively. In this illustrative case, the temperature of 15 °C has a maximum membership grade, amongst all CoLs, for cold, and hence, 15 °C is assigned with the CoL of cold. Referring back to (1), for computing the conditional relative frequency for cold, the numerator is going to sum all the data instances where the crisp inputs are assigned with cold for a given time point \(t_n\) such as a particular month of a year. The denominator finds the mode of occurrence of cold across all months. The result of the division will scale the \(g_{\text{cold}}\) values to [0, 1].

An illustration for calculating the \(g_{\text{cold}}\) values using (1), with a total of 12 time points as the months of a year is shown in Fig. 4(b) with continuous values of \(f_{\text{cold}}\), found using interpolation of \(g_{\text{cold}}\) plotted in Fig. 4(c). Please note the associated time intervals (as listed in the illustration in Fig. 4 are seasons in a year, such as winter, spring, summer, and autumn), are for easing the computational complexity of the four-dimensional (4-D) TT2FSs as will be explained later in Section III-D by taking time interval-based slice of the TT2FS.

### C. Temporal Type-2 Fuzzy Sets

In this section, a formal definition of TT2FS is presented. TT2FSs are 4-D as they incorporate information from the input domain \((X, T)\), frequency of occurrence domain \((F)\), and are characterized by a temporal membership function (TMF).

The computation of TMF, hereby termed as temporal fuzzification, involves the following two stages: 1) fuzzification of crisp input values of \(A\) from feature domain \(X\) to form \(T1\) \(\mu_A(x)\), as undertaken in standard T1 fuzzy sets; and 2) computation of the conditional distribution of \(A, f_A\). The temporal fuzzification is illustrated in Fig. 4(a) and defined as follows.

**Definition III.2 (TMF):** The TMF can be defined as

\[
\mu_A(x, t, f_A) = \mu_A(x) \otimes f_A
\]

where \(\otimes\) is a relation operator, \(\mu_A(x)\) is the primary membership of \(A\) in feature domain credited by the conditional distribution of \(A\), denoted \(f_A\), using the Mamdani relation (outlined earlier in Section III-A).

**Theorem III.1:** The TMF of \(A\), constructed using Mamdani relation (2), \(\mu_A(x, t, f_A)\) is \([0, 1]\).

**Proof:** The range of \(\mu_A(x, t, f_A)\) follows directly from the range of primary MF of \(A\): \(\mu_A(x)\subseteq[0, 1]\), and the conditional distribution of \(A\): \(f_A\subseteq[0, 1]\). Hence, by crediting \(\mu_A(x)\) with \(f_A\) using Mamdani relation (taking the min or product), it follows that the range of \(\mu_A(x, t, f_A)\subseteq[0, 1]\).

**Proposition III.1.1:** If the primary membership of TMF is normal and the conditional distribution \(f\) is normal, according to (1), then the resultant TMF membership after applying the Mamdani relation yields a normal TMF, therefore, we can imply that

\[
\sup_{x \in X} \mu_A(x, t, f_A) = 1.
\]

**Proof:** Given a \(f_A\subseteq[0, 1]\) and a \(\mu_A(x)\subseteq[0, 1]\) both with \(sup = 1\), \(\forall x \in X\) by deduction, \(\exists x : f_A \times \mu_A(x) \lor \min(f_A, \mu_A(x)) = 1\)

Next, we define the TT2FS, which are characterized by a TMF.

**Definition III.3 (TT2FS):** A TT2FS \(\tilde{A}\) of the universe of discourse \(X \times T \times F\) is characterized by a credited TMF \(\mu_A(x, t, f_A) : X \times T \times F \to [0, 1]\) where \(X\) is the feature domain of \(A\) characterized by a \(T1\) MF \(\mu_A(x)\), \(T\) is the time domain of \(A, F\) is the frequency of occurrence domain of \(A\) characterized by conditional frequency distribution with respect to time \(f_A\).
In mathematical set notation, $\tilde{A}$ can be written as

$$\tilde{A} = \{(x, t, f, \mu_\tilde{x}(x, t, f)) \mid \forall x \in X \quad \forall t \in T \quad \forall \mu_\tilde{x}(x, t, f) \subseteq [0, 1] \}$$

where $\mu_\tilde{x}(x, t, f) \subseteq [0, 1]$. Please note the conditional distribution, $f_A$, is a continuous distribution interpolated from discrete conditional relative frequency, $f_{x,t}$, and is defined mathematically earlier in (1). $\tilde{A}$ can also be expressed as

$$\tilde{A} = \int_{x \in X} \int_{t \in T} \int_{f \in F} \mu_\tilde{x}(x, t, f)/f_A/t/x$$

where $\int \int \int$ denotes the aggregation over all admissible values of $x$, $t$, and $f_A$. The associated TMF, $\mu_\tilde{x}(x, t, f_A) \subseteq [0, 1]$, scales the $\mu_A(x)$ based on its conditional distribution $f_A$ as defined in (2).

### D. Operations on TT2FSs

In this section, the common operations for TT2FSs such as the union and intersection, as well as defuzzification are outlined. TT2FSs, on account of being 4-D, are more computationally intense than GT2 fuzzy sets, which are 3-D. A popular approach for minimizing the computational demand of 3-D GT2 fuzzy sets is to use z-slice-based framework [16]. Motivated from the effectiveness of z-slice-based framework for simplifying the computations for GT2 fuzzy sets, in this work, the approach of taking time interval slice followed by z-slice (TS-ZS) is taken for performing operations on TT2FSs. The TS-ZS approach is explained in more detail as follows.

i) **TS:** Time interval-based slice to convert 4-D TT2FSs into 3-D. The 3-D time interval-based TT2FSs is similar to 3-D GT2 fuzzy set, with both sharing the feature domain on $x$-axis. On $y$-axis is the frequency of occurrence domain, for that time interval, for time interval-based TT2FS, while for GT2 fuzzy sets, primary membership grade is on $y$-axis. And on $z$-axis is the temporal membership grade for time interval-based TT2FS while for GT2 fuzzy set secondary membership grade is on $z$-axis.

ii) **ZS:** z-slice-based approach for the time interval-based 3-D TT2FS as utilized for GT2 fuzzy sets. The z-slices at specific $z$-levels render a given 3-D fuzzy set to an equivalent IT2 fuzzy set with lower and upper primary membership grades. For the case of TS-ZS-based TT2FSs, the primary membership grades are the conditional distribution values for that time interval at a given $z$-level.

In the following sections, a formal definition for the operations on TT2FSs is given with $\tilde{A}$ and $\tilde{B}$ denoting two TT2FSs characterized by TMFs $\mu_\tilde{x}(x, t, f_A)$ and $\mu_\tilde{y}(x, t, f_B)$, respectively, as outlined in the following equation:

$$\tilde{A} = \int_{x \in X} \int_{t \in T} \int_{f \in F} \mu_\tilde{x}(x, t, f)/f_A/t/x$$

$$\tilde{B} = \int_{x \in X} \int_{t \in T} \int_{f \in F} \mu_\tilde{y}(x, t, f)/f_B/t/x$$

where $X$ is the feature domain, $T$ is the time domain, and $F$ is the frequency of the occurrence domain.

1) **Union and Intersection Operations:** A general procedure for undertaking the union and intersection operations on the 4-D TMFs is outlined in Algorithm 1. The union of two TT2FSs $\tilde{A}$ and $\tilde{B}$ is a TT2FS defined as $\tilde{A} \cup \tilde{B}$ in the following equation:

$$\tilde{A} \cup \tilde{B} = \int_{x \in X} \int_{t \in T} \int_{f \in F} \mu_{\tilde{A} \cup \tilde{B}}(x, t, f)/f/t/x$$

where $\mu_{\tilde{A} \cup \tilde{B}}$ can be calculated by discretizing the $T$ domain, and taking z-slices on $\mu_{\tilde{A} \cup \tilde{B}}$. The $z$-slices as outlined in (Alg 1.1) of Algorithm 1. In particular, for union operation, at time interval $\Delta t_q$ (Alg 1.1) takes the form of (8) when using the max $t$-conorm

$$\mu_{\tilde{A} \cup \tilde{B}}(x, t, f_{\Delta t_q}) = \sum_x \sum_{f_{\Delta t_q} \leq \max(t_A, t_B), \max(u_A, u_B)} z_i/f_{\Delta t_q}$$

Likewise, the intersection of TT2FSs can be written as shown in the following equation:

$$\tilde{A} \cap \tilde{B} = \int_{x \in X} \int_{t \in T} \int_{f \in F} \mu_{\tilde{A} \cap \tilde{B}}(x, t, f)/f/t/x$$
where $\mu_{\tilde{A}_{t_i}}$ can be calculated by discretizing the $T$ domain, and taking $z$-slices on $\mu_{\tilde{A}_{t_i}}(x,t,f)$ values as outlined in (Alg 1.1) of Algorithm 1. In particular, for intersection operation, at time interval $\Delta t_q$ (Alg 1.1) takes the form of (10) when using the min $t$-norm. However, please note either product or min can be applied

$$\mu_{\tilde{A}_{t_i} B_{t_i}}(x, f_{\Delta t_q}) = \sum_{x} \sum_{f_{\Delta t_q} \in [\min(I_{x},I_{y}), \min(u_{y},u_{b})]} z_i / f_{\Delta t_q}. \tag{10}$$

2) Defuzzification: In general, defuzzification converts a fuzzy set to an equivalent crisp number, and can be thought of as the inverse of fuzzification. For T1 fuzzy sets, defuzzification usually involves computing the centroid of the T1 fuzzy set [17] to compute a representative crisp number, as shown in the following equation:

$$x^* = \frac{\sum_{b=1}^{B} x_b \mu(x_b)}{\sum_{b=1}^{B} \mu(x_b)} \tag{11}$$

where $x^*$ is the centroid of the T1 MF defined on the domain $x \in X$. Here, the summation sign is used as in typical mathematical equations, i.e., for the case of the numerator, it is summing the product of $x$ values and their corresponding membership values whereas for the denominator it is summing the membership values corresponding to all $x$ values $\forall b \in [1, \ldots, B]$.

For a 3-D GT2 fuzzy set, defuzzification usually involves the following three steps, outlined as follows.

i) Transforming a 3-D GT2 fuzzy set to IT2 fuzzy sets by slicing the GT2 fuzzy set at given $z$-levels such as $z_i \in [z_1, \ldots, z_l]$.

ii) Type reducing the $z$-level-based IT2 fuzzy sets results in two T1 fuzzy sets using the Karnik–Mendel (KM) method [18]. The type-reduced T1 fuzzy sets are composed of the left and right centroids of the IT2 fuzzy sets. More specifically, the KM method requires iterative process to compute left and right centroids resulting in two T1 fuzzy sets: $[y_{l_{z_1}}, y_{l_{z_2}}, \ldots, y_{l_{z_l}}]$ and $[y_{r_{z_1}}, y_{r_{z_2}}, \ldots, y_{r_{z_l}}]$ where $y_{l_{z_i}}$ is the left centroid at $z$-level 1 and $y_{r_{z_i}}$ is the right centroid at $z$-level 1 and so on.

iii) Defuzzification of the type reduced T1 fuzzy sets, using centroid average, to find equivalent $y_{l_{\Delta t_q}}$ and $y_{r_{\Delta t_q}}$:

$$y_{l_{\Delta t_q}} = \frac{(z_1 * y_{l_{z_1}}) + (z_2 * y_{l_{z_2}}) + \cdots + (z_l * y_{l_{z_l}})}{z_1 + z_2 + \cdots + z_l} \tag{12}$$

$$y_{r_{\Delta t_q}} = \frac{(z_1 * y_{r_{z_1}}) + (z_2 * y_{r_{z_2}}) + \cdots + (z_l * y_{r_{z_l}})}{z_1 + z_2 + \cdots + z_l} \tag{13}$$

iv) The final type-reduced crisp value is found using the Nie–Tan method [19] on $y_{l_{\Delta t_q}}$ and $y_{r_{\Delta t_q}}$.

Algorithm 2: Defuzzification of TT2FSs for a Given Time Interval $\Delta t_q$.

Result: Crisp value for a given time interval, denoted by $\text{crisp}_{\Delta t_q}$, where $\Delta t_q$ is the $q$th time interval.

Let feature $A$ on feature domain $X$ have temporal membership function (TMF) denoted by $\mu_A(x,t,f_k(t,\mu_A(x)))$ with time intervals $\Delta t_q \in [\Delta t_1, \ldots, \Delta t_Q]$ and $z$-slices discretised at $z_i \in [z_1, z_2, \ldots, z_l]$;

For each 3D time interval based TMF, the defuzzification can be done independently, by first taking the $z$-slices at $z_i \in [z_1, z_2, \ldots, z_l]$ which renders the 3D time interval based TMF into interval type 2 (TT2) MFs;

The left and right centroid for each IT2 TMF at $z$-location $z_i$, denoted by $C_{z_i,\Delta t_q}$, can be computed using Karnik-Mendel (KM) method [18] to give $[y_{l_{z_i}}, y_{r_{z_i}}]$ at that $z$-slice $z_i$ and time interval $\Delta t_q$, as outlined in eq. (Alg 2.1);

for $z_i \leq z_t$ do

$$C_{z_i,\Delta t_q} = [y_{l_{z_i,\Delta t_q}}, y_{r_{z_i,\Delta t_q}}] \tag{Alg 2.1}$$

end

Defuzzification of the type reduced T1 fuzzy sets, using centroid average, to find equivalent $y_{l_{\Delta t_q}}$ and $y_{r_{\Delta t_q}}$:

$$y_{l_{\Delta t_q}} = \frac{(z_1 * y_{l_{z_1,\Delta t_q}}) + (z_2 * y_{l_{z_2,\Delta t_q}}) + \cdots + (z_l * y_{l_{z_l,\Delta t_q}})}{z_1 + z_2 + \cdots + z_l} \tag{Alg 2.2}$$

$$y_{r_{\Delta t_q}} = \frac{(z_1 * y_{r_{z_1,\Delta t_q}}) + (z_2 * y_{r_{z_2,\Delta t_q}}) + \cdots + (z_l * y_{r_{z_l,\Delta t_q}})}{z_1 + z_2 + \cdots + z_l} \tag{Alg 2.3}$$

A crisp value, $\text{crisp}_{\Delta t_q}$, can now be computed by applying Nie-Tan method [19] on $y_{l_{\Delta t_q}}$ and $y_{r_{\Delta t_q}}$.

In this work, the defuzzification of 4-D TT2FS also involves TS-ZS approach (explained earlier in Section III-D), i.e., taking the time interval-based slice followed by $z$-slices. The time interval-based TMF is 3-D, and for each of the time interval ($\Delta t_q$) based TMF, $z$-slices at particular $z_i$ levels renders them as IT2 fuzzy sets. The KM procedure [18] can be applied on IT2 fuzzy sets, at each $z$-level, to compute T1 fuzzy sets composed of $[y_{l_{z_1,\Delta t_q}}, y_{r_{z_1,\Delta t_q}}]$ as outlined in (Alg 2.1). Using the centroid defuzzifier, the T1 fuzzy sets are defuzzified to give one equivalent $y_t$ and $y_r$, for that time interval, as outlined in (Alg 2.2) and (Alg 2.3). The Nie–Tan method [19] is then applied to compute one crisp value for that time interval. The defuzzification of TT2FSs, for a given time interval, is summarized in Algorithm 2. The procedure outlined in Algorithm 2 can be repeated for each time interval, i.e., $\Delta t_q$ where $q \in [1, \ldots, Q]$, to obtain a crisp value for all time intervals.

IV. TXAI INFERENCE SYSTEM

In this section, the TXAI-IS for classification problems is outlined. A general flowchart for the TXAI-IS is outlined in Fig. 5. The temporal fuzzifier constructs the 4-D TT2FSs, as outlined in Fig. 4(a). To analyze a given dynamic process with...
Fig. 5. General schematic representation delineating the interlinks between salient components of a TXAI-IS.

respect to time, the TXAI-IS works for each time interval $\Delta t_q$ where $\Delta t_q \in [\Delta t_1, \ldots, \Delta t_Q]$ independently. To this end, the 4-D TT2FSs are first sliced based on the $\Delta t_q$, and inference is made on time sliced 3-D TT2FSs using the temporal rules for the same $\Delta t_q$. Each time interval would entail a unique temporal rule base. The temporal rules can either be furnished by experts in the field or can be learnt from the input data using evolutionary algorithms, such as genetic algorithm (GA) [20].

In addition, the assumptions of the proposed TXAI system with TT2FSs include the following:
1) the input features and output are observable;
2) a relation between input features and output exists;
3) the relation between input features and output varies with time.

In the following sections, the classification TXAI-IS is outlined in detail as the empirical study on which TXAI system is exemplified also undertakes a classification problem, i.e., occupancy dataset [11] is analysed to determine whether or not a room is occupied.

A. Classification

For the classification problem, the TXAI-IS will predict one class or label for a given data instance for each time interval. The overall TXAI-IS for classification undertakes the following steps.

i) Compute the membership degree for the time interval-based 3-D TT2FSs.
   a) The time interval-based 3-D TT2FSs are transformed into IT2 fuzzy sets by taking slices at predefined $z$-levels. The degree of membership at each $z$-level, such as $z_i \in [z_1 \ldots, z_J]$ where $J$ is the total number of $z$-slices, for a given 3-D TT2FS $A$ is given as follows [16]:
   \[
   \tilde{A} = \{(x, u, z) | \forall x \in X \forall u \in [\mu_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)] \subseteq [0, 1]\}
   \]
   where $\mu_{\tilde{A}}$ is the membership degree of the IT2 fuzzy set $\tilde{A}$ at the predefined $z$-level.
   ii) Compute the firing strength for each rule, at each $z$-level.
      a) The upper and lower firing strength of a given rule $p$, $w_p$, and $\bar{w}_p$, respectively, is the degree of match between the rule $p$ and the data instance $x$. It is computed as
      \[
      w_p(x^k) = \prod_{k=1}^{a} \tilde{A}(x^k)
      \]
      \[
      \bar{w}_p(x^k) = \prod_{k=1}^{a} \bar{\mu}_{\tilde{A}}(x^k)
      \]
      where $p$ is the rule number, $a$ is the total number of antecedents in the rule $p$ and $x^k$ is an input $(k)$ of the actual $k$-dimensional data instance to be classified. We recommend scaling the bounds $(w_p(x))$ and $(\bar{w}_p(x))$ of a given rule $(p)$, dividing each bound by its respective sum of firing strengths (either lower or upper) of all the rules in the rule-base having the same consequence as $(p)$.
   iii) Compute the rule weight (RW) for each rule, at each $z$-level.
      a) The RW is a measure of a given rule’s dominance and is computed as
      \[
      \text{RW}_p = \tau_p \times \bar{\tau}_p
      \]
      \[
      \text{RW}_p = \zeta_p \times \bar{\zeta}_p
      \]
where $c$ is the confidence of the rule $p$ and $s$ is the support of the $p$th rule.

b) The confidence of a rule is a measure of the likelihood to correctly classify a given data instance. It is calculated as shown in (17)

$$
\bar{c}_p(Ants_p \Rightarrow Cons_p) = \frac{\sum_{x \in (Ants_p \Rightarrow Cons_p)} \Pi_p(x)}{\sum_{p=1}^{P} \sum_{x \in (Ants_p \Rightarrow Cons_p)} \Pi_p(x)}
$$

$$
\bar{s}_p(Ants_p \Rightarrow Cons_p) = \frac{\sum_{x \in (Ants_p \Rightarrow Cons_p)} w_p(x)}{\sum_{p=1}^{P} \sum_{x \in (Ants_p \Rightarrow Cons_p)} w_p(x)}
$$

(17)

where $Ants_p$ and $Cons_p$ are the antecedents and consequent, respectively, of the rule $p$. The numerator sums the firing strength of all the data instances that have the same antecedents and consequent as the rule $p$. Whereas the denominator sums the firing strength of all the data instances that have the same antecedents as the rule $p$ irrespective of the consequent—for all the rules $[1, \ldots, P]$, where $P$ is the total number of rules.

c) The support of a rule is calculated as shown in (18)

$$
\bar{s}_p(Ants_p \Rightarrow Cons_p) = \frac{\sum_{x \in (Ants_p \Rightarrow Cons_p)} w_p(x)}{P}
$$

$$
\bar{c}_p(Ants_p \Rightarrow Cons_p) = \frac{\sum_{x \in (Ants_p \Rightarrow Cons_p)} \Pi_p(x)}{P}
$$

(18)

with $P$ as the total number of rules.

d) Compute the association degree of each rule, with a given data instance, for each z-level.

a) The association degree of a rule $p$ with a given data instance $x$ is computed as

$$
\bar{h}_p = \Pi_p(x) \times \bar{R}W_p
$$

$$
\bar{h}_p = \Pi_p(x) \times \bar{R}W_p.
$$

(19)

e) Predict the label.

i) Find a value of the association degree, $h$, for each rule by using the Nie–Tan [19] method on the $h$ and $\bar{h}$, which are found using (Alg 2.2) and (Alg 2.3).

b) The rule with the highest association degree, $h$, predicts the label for the given data instance.

c) The steps outlined above (i)–(v) are repeated for each time interval to predict a label for all time intervals.

### B. Numerical Step-Wise Example

In this section, a binary classification problem using TXAI-IS is exemplified using a hypothetical dataset with two input features, Feature 1 and Feature 2, and one output. Let time intervals be defined over a day, such as morning, daytime, and evening with three CoLs associated with the inputs (Feature 1 and Feature 2) be: [low, medium, high] and output labels be Output 1 and Output 2. First, TTFSs for both inputs (Feature 1 and Feature 2) are constructed using a temporal fuzzifier, as outlined in Fig. 4. Also, for each time interval, the rules will be different but the overall process to determine the output label is same. In the following steps, we exemplify how the output label is predicted for one time interval, in this example, morning.

#### Table II

| CoLs | z_{0.2} | z_{0.4} | z_{0.6} | z_{0.8} | z_{1.0} |
|------|---------|---------|---------|---------|---------|
| Low  | 0.50    | 0.52    | 0.54    | 0.52    | 0.51    |
| Med. | 0.61    | 0.63    | 0.64    | 0.61    | 0.60    |
| High | 0.63    | 0.63    | 0.65    | 0.63    | 0.61    |

The association degrees, crisp values, for each of the rules $R_1$–$R_3$, denoted $h$ is also listed.

Let the rules ($R$) outlining the relation between input features and output for morning be as listed in (20). The corresponding lower and upper RW at each $z$-level are as listed in Table III. In the following steps (i)–(iv), we show how a corresponding label for output is predicted using TXAI-IS for input values of Feature 1 = 19.7 and Feature 2 be = 4.3. In this example, the $z$-level is discretised at $z_{0.2}$, $z_{0.4}$, $z_{0.6}$, $z_{0.8}$, and $z_{1.0}$.

$R_1$: IF Feature 1 is Low and Feature 2 is Medium

**THEN** Output is Output 2

$R_2$: IF Feature 1 is Medium and Feature 2 is Medium

**THEN** Output is Output 1

$R_3$: IF Feature 1 is High and Feature 2 is High

**THEN** Output is Output 1

(20)

i) The degree of membership for each CoL of the inputs Feature 1 and Feature 2 is determined from the time interval (Morning) based 3-D TMF. The membership degree is the value of the conditional distribution at a given input value and corresponding $z$-level as outlined in (14). Let the corresponding membership degrees for each CoL of the inputs Feature 1 and Feature 2 be as noted in Table II.

ii) The firing strength of each rule listed in (20) are found, using the membership degree in Table II, as outlined in (15) and listed in Table III. As an example, for $R_1$, the lower firing strength at $z = 0.6$, $w_{1,z=0.6}^\alpha$, can be calculated as follows:

$$
w_{1,z=0.6}^\alpha(x = [19.7, 4.3]) = \prod_{k=1}^{2} \mu(x^k)
$$

$$
= 0.54 \times 0.55 = 0.297.
$$

(21)

iii) The association degree of each rule with the input data instance is determined, using the firing strength in Table III, as outlined in (19). The upper and lower values of the association degree for the five $z$-levels are as listed.
TABLE III
LOWER AND UPPER FIRING STRENGTHS, \( w \) AND \( \pi \), RESPECTIVELY, FOR THE HYPOTHETICAL RULES LISTED IN (20) FOR TIME INTERVAL MORNING

| Rule | Firing Strength, \( w \) | \( z_{0.2} \) | \( z_{0.4} \) | \( z_{0.6} \) | \( z_{0.8} \) | \( z_{1.0} \) | Consequent | \( w \) | \( z_{0.2} \) | \( z_{0.4} \) | \( z_{0.6} \) | \( z_{0.8} \) | \( z_{1.0} \) |
|------|---------------------|-------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|-------|
| \( R_1 \) | Lower | 0.25 | 0.286 | 0.297 | 0.281 | 0.27 | output 2 | Lower | 0.31 | 0.30 | 0.30 | 0.29 | 0.27 |
|       | Upper | 0.354 | 0.372 | 0.378 | 0.354 | 0.342 |            | Upper | 0.35 | 0.34 | 0.34 | 0.31 | 0.30 |
| \( R_2 \) | Lower | 0.315 | 0.347 | 0.358 | 0.34 | 0.323 | output 1 | Lower | 0.69 | 0.69 | 0.68 | 0.66 | 0.66 |
|       | Upper | 0.447 | 0.46 | 0.46 | 0.447 | 0.427 |            | Upper | 0.73 | 0.73 | 0.72 | 0.72 | 0.72 |
| \( R_3 \) | Lower | 0.26 | 0.256 | 0.256 | 0.265 | 0.277 | output 1 | Lower | 0.22 | 0.21 | 0.21 | 0.21 | 0.21 |
|       | Upper | 0.297 | 0.297 | 0.313 | 0.313 | 0.328 |            | Upper | 0.24 | 0.22 | 0.22 | 0.22 | 0.22 |

The RWS at each \( z \)-level are also listed.

The association degrees’ crisp value, for each of the rules \( R_1 - R_3 \), denoted \( h_{\text{crisp}} \) is also listed.

in Table IV. As an example, for \( R_2 \) the upper association degree at \( z = 0.2 \), \( \bar{h}_{2z=0.2} \), can be calculated as follows:

\[
\bar{h}_{2z=0.2} = \overline{w_{z=0.2}}(x) \times \overline{\text{RW}_{2z=0.2}} = 0.447 \times 0.73 = 0.326. \tag{22}
\]

iv) The consequent of the rule with the highest association degree with the input data instance becomes the predicted label for a given time interval. The crisp value for the association degree of each rule is found using (Alg 2.2) and (Alg 2.3). As an example, the crisp value of association degree for \( R_3 \) is found as follows:

\[
h_{3_1} = \frac{0.2 \times (h_{3_{0.2}}) + \ldots + 1.0 \times (h_{3_{1.0}})}{0.2 + 0.4 + 0.6 + 0.8 + 1.0} = \frac{0.2 \times 0.057 + 0.4 \times 0.054 + \ldots + 1 \times 0.058}{3} = 0.056
\]

\[
h_{3_2} = \frac{0.2 \times (h_{3_{0.2}}) + \ldots + 1.0 \times (h_{3_{1.0}})}{0.2 + 0.4 + 0.6 + 0.8 + 1.0} = \frac{0.2 \times 0.071 + 0.4 \times 0.065 + \ldots + 1 \times 0.072}{3} = 0.0696
\]

\[
h_{3_{\text{upper}}} = \frac{0.056 + 0.0696}{2} = \frac{0.1256}{2} = 0.063. \tag{23}
\]

In this illustrative example, \( R_2 \) has the highest association degree (tabulated in Table IV), hence, the predicted output for the input data instance (Feature 1 = 19.7 and Feature 2 = 4.3) for time interval Morning is the consequent of \( R_2 \), i.e., Output 1.

The same process can be repeated for each time interval with their respective rules to predict a label for the output. Hence, in this numerical example, there will be three output labels for a total of three time intervals.

C. Estimating Temporal Trajectories From TXAI Models

The temporal trajectories of a dynamic system can be outlined by the TXAI system by making use of the conditional distribution integrated into the TXAI system. The trajectories of a TXAI model is motivated by the work of Filev et al. [21] that embodies fuzzy transition events defined by joint possibility encompassing the current and future prototypical rules. More specifically, the TXAI system can delineate a rule transition matrix (RTM), which will entail the joint possibility of the rules in present \((\Delta t)\) and future \((\Delta t^+\)) time intervals. In mathematical terms, for a total of \( U \) rules in time interval \( \Delta t \), and a total of \( V \) rules in time interval \( \Delta t^+ \), the RTM can be written as follows [21]:

\[
\text{RTM}(\Delta t, \Delta t^+) = \begin{bmatrix}
\pi_{11} & \ldots & \pi_{1N} \\
\vdots & \ddots & \vdots \\
\pi_{1M1} & \ldots & \pi_{1MV}
\end{bmatrix} \tag{24}
\]

where \( \pi_{cd} \) is the rule transition possibility (RTP) for the \( c \)-th rule, \( r_c \), in time interval \( \Delta t \) and the \( d \)-th rule, \( r_d \), in time interval \( \Delta t^+ \) as given by the following equation:

\[
\pi_{cd} = \eta_{cd} \times \frac{S_{cd}}{S_{\Delta t^+}} \tag{25}
\]

where \( \eta_{cd} \) is the joint possibility for the two rules to be prototypical in their respective time intervals, and the ratio \( \frac{S_{cd}}{S_{\Delta t^+}} \) entails the number of times \( r_c \) and \( r_d \) are observed in their respective time intervals with respect to all \( V \) rules in \( \Delta t^+ \). The following (26)–(28), outline how \( \eta_{cd} \) and the ratio \( \frac{S_{cd}}{S_{\Delta t^+}} \) are computed

\[
\eta_{cd}(r_c, \Delta t, r_d, \Delta t^+) = \gamma(c(r_c, \Delta t) \times \gamma(d(r_d, \Delta t^+)) \tag{26}
\]

where \( \gamma \) is computed by applying the \( t \)-norm operator (product or minimum type) to the conditional distribution values of the antecedents of a given rule \( r \) in a given time interval \( (\Delta t) \) or \( (\Delta t^+) \); mathematically expressed as shown in equation (27) for rule \( (r_c) \) in time interval \( (\Delta t) \). The computation of the conditional distribution, \( f \), is previously outlined in Section II-C [in
particular see (1)]
\[
\gamma_c(r_c, \Delta t) = f_c(\text{Ant}_1, r_c, \Delta t) \times f_c(\text{Ant}_2, r_c, \Delta t) \\
\times \cdots \times f_c(\text{Ant}_n, r_c, \Delta t)
\]
(27)
where \( r_c \) is the total number of antecedents (Ant) of rule \( r_c \). The elements for computing the ratio \( \frac{\text{Numerator}}{\text{Denominator}} \) are outlined in (28)
\[
S_{\text{cd}} = \sum r_{c, \Delta t} r_{d, \Delta t}^+ \\
S_{\Delta t^+} = \sum_{d=1}^{V} r_{d, \Delta t^+}
\]
(28)
where the numerator, \( S_{\text{cd}} \), represents the sigma count of the number of times \( r_c \) and \( r_d \) are observed in their respective time intervals, and the denominator, \( S_{\Delta t^+} \), denotes the sigma count of observing all \( V \) rules in \( \Delta t^+ \).

V. CASE STUDY: TIME-DEPENDENT OCCUPANCY DATASET

In this section, a temporal occupancy dataset [11] is used to exemplify the proposed TXAI system modeling. The occupancy dataset entails measurements of a room along with the time of when the measurement is recorded. In particular, it includes measurements of the room temperature, light, CO\(_2\), and a binary label of whether or not the room is occupied. There are 8143 data instances in the dataset taken over a period of a few weeks.

In this work, the dataset [11] is used for classification problem where TXAI system predicts whether or not the room is occupied based on the room measurements. The inputs of temperature, light, and CO\(_2\) are used to predict whether or not the room is occupied. Three CoLs of low, medium, and high are associated with inputs of temperature, light, and CO\(_2\). The primary MF of the CoLs for all inputs are empirically found. The time is discretized at each hour of the day, hence, a total of \( N = 24 \) time points with a total of three time intervals defined at morning, daytime, and evening, as also summarized in Table V. The \( z \)-slices are obtained on locations [0.2, 0.4, 0.6, 0.8, 1.0]. All aforementioned parameters values are selected so as to reflect the inherent dynamics of the system (such as discretizing time at each hour) and to obtain a good enough TXAI model without adding too much computational complexity, for example, the more the \( z \)-slices the more accurate the TXAI model would be but at a greater computational cost (\( \rho = Q \times z_j \) but independent of the data size in each \( \Delta t_{ij} \)).

The conditional distribution for each CoL of every input is computed on the entire dataset. Once the conditional distributions are computed, the learning procedure focuses on the data belonging to each interval. A 10-repeated nested cross-validation procedure is adopted. The dataset is split into a disjoint stratified train, validation, and the test set to ensure a random selection of the datasets (train, validation, and test) is not creating any bias in the results. Each repetition, 20\% of the dataset is held out as a test set, and the remaining is used to build the train and validation sets. Train and validation sets are determined in an inner 10-fold procedure, where a fold is used for validation and the rest for training to determine the RWs. Balanced accuracy and other performance metrics are computed over each validation and test set.

A rule-base is formed for each time interval. The rules are learned using GA [22] such that they (rules) attain optimally balanced accuracy on the validation datasets. The GA parameters specification includes the number of generations, set at 20, with each generation having a population of 50. Moreover, the GA is leveraged to find the rules that are prototypical for each time interval. The number of antecedents in each rule can be at most 3 but not more to underpin explainability and hamper model complexity, therefore precluding overfitting. For the same reason, the maximum number of rules in each candidate rule-base for each time interval was limited to 30, although further pruned when its weight [see (16)] does not surpass a tolerance threshold of 0.001.

In order to compare the performance of the proposed TXAI system, numerous state-of-the-art classifiers which can both analyze time-series data and/or are explainable have been used. More specifically, for comparison with temporal analysis long short-term memory (LSTM) [23] and hidden Markov models (HMM) [24] are used, for comparison with explainable models the standard GT2-based XAI system is used, and for partial explainability DT [25] is used. In addition, a comparison is also made with a temporal convolutional network (TCN) [26] for comparison with deep learning methods [27]. Parametrization and configuration was set to default mode of their respective libraries (Sklearn and Keras). For methods with no modeling with respect to a time component, time is given as an extra input feature. Moreover, the train, validation, and test dataset splits are similar across all methods and for GT2-based XAI in particular, the location of \( z \)-slices, and the GA parameters for rule learning are also identical to those of TXAI system.

| Problem            | Input/output | Feature/label | CoLs                  | N  | Time Intervals, \( \Delta t \)     |
|--------------------|--------------|---------------|-----------------------|----|-----------------------------------|
| Classification     | Input        | Temperature   | Low, medium, high     | 24 | Morning, daytime, evening         |
|                    |              | Light         | Low, medium, high     | 24 | Morning, daytime, evening         |
|                    |              | CO\(_2\)      | Low, medium, high     | 24 | Morning, daytime, evening         |
| Output             | Occupied     |               |                       | -  | Morning, daytime, evening         |
|                    | not occupied |               |                       | -  |                                   |

The output for the classification problem predicts the label of whether the room is occupied or not. The time points \( t_c \) for calculating the frequency of occurrence are 24 on account of the number of hours in a given day with a total of three corresponding time intervals \( \Delta t_i \) of morning (Time < 11 AM), daytime (11 AM < Time < 7 PM), and evening (Time > 7 PM).

### Table V

Classification Problem is Exemplified Using the Proposed TXAI System With Occupancy Dataset [11]
A. Results

For the classification problem undertaken, using the occupancy dataset, the proposed TXAI system, and numerous state-of-the-art classification methods predict whether or not the room is occupied. The mean (and standard deviation) f-score obtained using TXAI system on the 10 test datasets is 95.30%, which is the highest score on the test dataset across all classifiers except TCN. The other classification metrics investigated in this work are balanced accuracy, recall, and precision. A bar plot for the aforementioned classification metrics for both the proposed TXAI and the state-of-the-art AI methods (TCN, LSTM, DT, HMM, GT2-based XAI) on 10 times repeated 10-fold validation and test datasets is shown in Fig. 6(a) and (b), respectively. In addition, a convergence graph that outlines how the GA optimisation converges with respect to balanced accuracy for both TXAI and GT2-based XAI systems is also shown in Fig. 6(c).

The rules outlined by TXAI and GT2-based XAI systems which are prototypical for whether or not the room is occupied are listed in Table VI. For the TXAI system, please note that the rules are found separately for each time interval (morning, daytime, and evening) whereas, for the GT2-based XAI system, the time intervals are one of the antecedents of the rules. In general, for both TXAI and GT2-based XAI systems, the rules outline that when the room measurements have higher values, the room is more likely to be occupied, and when the room measurements are on the lower end, the room is more likely to be not occupied.

For the TXAI system, the temporal trajectories of a time-variant system can also be investigated using the RTMs, previously outlined in Section IV-C. The individual RTMs transitioning from one time interval ($\Delta t$) to another, i.e., from morning to daytime, from daytime to evening, and from evening to morning, represent the joint possibilities of observing a given rule in $\Delta t^+$ with respect to the rules in $\Delta t$. The rules corresponding to the highest rule transition possibilities (RTPs) are also joined with lines in the column rule transitions (RTs) in Table VI and illustrated in a schematic in Fig. 7.

B. Discussion

In this work, the proposed TXAI system is used to model an occupancy dataset [11] for the classification problem of whether or not the room is occupied. For comparison purposes, several state-of-the-art explainable (GT2-based XAI system), partially explainable (DT), and nonexplainable methods that can analyse temporal information (LSTM and HMM) as well as TCN are also applied to the aforementioned classification problem. As can be noted from the Fig. 6(a) and (b), TXAI offers greater classification performance than all classifiers (for, e.g., for mean f-score TXAI performs better than LSTM by 18.19%, DT by 6.81%, HMM by 4.90%, GT2-based XAI system by 8.58% on test datasets) except TCN (for mean f-score TXAI performs better than GT2-based XAI system by 4.67% on test datasets). However, the TCN classification mechanism is not explainable hence unable to shed light on the prediction of the room occupancy based on input features of Temperature, Light, $CO_2$, and Time.

With respect to the comparison with the GT2-based XAI system, the only explainable system apart from the proposed TXAI system, a convergence graph plotted in Fig. 6(c) also highlights that TXAI system converges ($\sim$500 function evaluations)
twice as faster than standard GT2-based XAI system (~1000 function evaluations) whilst also yielding higher classification metrics [see Fig. 6(a) and (b)]. Moreover, the rules outlined by the explainable systems, TXAI and XAI systems, are listed in Table VI, and both systems are in agreement that when the room measurements (temperature, light, and CO$_2$) have higher values, then the room is likely to be occupied, and when the room measurements are lower, then the room is likely to be not occupied. However, the rules for TXAI also offer greater insight into how the room measurements are interlinked with respect to predicting room occupancy. For example, for the time interval morning, rule no 5 (see Table VI) outlines that if both inputs of temperature and light have high values then the room is likely to be occupied. In this regard, rules across time intervals shed light on the intertwined CoLs of the inputs prototypical for decoding the room occupancy.

Furthermore, the TXAI systems are also able to shed light on the temporal trajectories of the system being modeled using RTMs, previously outlined in Section IV-C, and illustrated in Fig. 7. The rule transition possibilities (RTPs), which are the elements of the RTMs, entail the likelihood of observing the transition of a real-life process from one time point to another. The proposed TXAI system can shed light not only on which rules are prototypical for each of the time intervals but also on the likelihood of observing the rules across the different time points.

### TABLE VI

| Method | Time  | Rule no. | Rule                                         | Rule Weight | Rule Transition [RT] |
|--------|-------|----------|----------------------------------------------|-------------|----------------------|
| TXAI   | Morning | 1        | IF Light is High THEN room is Occupied      | 0.346       |                      |
|        |        | 2        | IF Temperature is High THEN room is Occupied | 0.099       |                      |
|        |        | 3        | IF CO$_2$ is Medium THEN room is Occupied   | 0.050       |                      |
|        |        | 4        | IF CO$_2$ is High THEN room is Occupied     | 0.046       |                      |
|        |        | 5        | IF Temperature is High AND Light is High THEN room is Occupied | 0.018       |                      |
|        |        | 6        | IF Light is High AND CO$_2$ is Medium THEN room is Occupied | 0.014       |                      |
|        |        | 7        | IF Light is High AND CO$_2$ is High THEN room is Occupied | 0.012       |                      |
|        |        | 8        | IF Temperature is Medium THEN room is Occupied | 0.013       |                      |
|        |        | 9        | IF Temperature is High AND CO$_2$ is High THEN room is Occupied | 0.011       |                      |
|        |        | 10       | IF Temperature is Medium AND CO$_2$ is Medium THEN room is Occupied | 0.007       |                      |
|        |        | 11       | IF Light is Low THEN room is Not Occupied   | 1.000       |                      |
|        |        | 12       | IF Temperature is Low THEN room is Not Occupied | 0.073       |                      |
|        | Daytime | 1        | IF Light is High THEN room is Occupied      | 0.473       |                      |
|        |        | 2        | IF Temperature is High THEN room is Occupied | 0.277       |                      |
|        |        | 3        | IF CO$_2$ is High THEN room is Occupied     | 0.110       |                      |
|        |        | 4        | IF Temperature is Medium AND Light is High THEN room is Occupied | 0.017       |                      |
|        |        | 5        | IF Temperature is High AND Light is High AND CO$_2$ is High THEN room is Occupied | 0.015       |                      |
|        |        | 6        | IF Light is Low THEN room is Not Occupied   | 1.000       |                      |
|        |        | 7        | IF CO$_2$ is Low THEN room is Not Occupied  | 0.50        |                      |
|        |        | 8        | IF Light is Medium THEN room is Not Occupied | 0.147       |                      |
|        |        | 9        | IF Temperature is High AND Light is Low THEN room is Not Occupied | 0.011       |                      |
|        |        | 10       | IF Light is High THEN room is Occupied      | 0.095       |                      |
|        |        | 11       | IF Light is Low THEN room is Not Occupied   | 1.000       |                      |
|        |        | 12       | IF Light is Low AND CO$_2$ is Low THEN room is Not Occupied | 0.104       |                      |
|        |        | 13       | IF Temperature is High AND Light is Low THEN room is Not Occupied | 0.041       |                      |
|        | Evening| 1        | IF Light is High AND Time is Daytime THEN room is Occupied | 0.540       |                      |
|        |        | 2        | IF Light is High AND Time is Morning THEN room is Occupied | 0.425       |                      |
|        |        | 3        | IF Temperature is High AND Time is Daytime THEN room is Occupied | 0.419       |                      |
|        |        | 4        | IF Light is Low AND Time is Morning THEN room is Not Occupied | 1.000       |                      |
|        |        | 5        | IF CO$_2$ is Medium AND Time is Evening THEN room is Not Occupied | 0.789       |                      |

Note: In the column RT, the rules with the highest rule transition possibility (RTP) for transitioning from one time interval to another are marked with connecting lines: red line connects the rules with the highest RTP for going from Morning to Daytime, blue line connects the rules with the highest RTP for going from Daytime to Evening, and green lines connect the rules with the highest RTP for going from Evening to Morning of the next day. The numerical values of the corresponding RTPs are also listed. The rules obtained using the standard general type-2 (GT2) explainable artificial intelligence (XAI) system with time as another input are also outlined at the end of the table for comparison purposes.
VI. CONCLUSION

The ability of an explainable system to model a real-life process in terms of its characteristic features is of paramount significance to inform about the nature of the process. In this regard, XAI systems have proved pivotal for increasing our understanding of numerous complex real-life processes. However, non-TXAI systems are not able to analyse real-life processes across time. This is a critical limitation of standard XAI systems for modeling time-variant real-life processes, especially where time is a defining parameter for the model, i.e., the real-life process behaves differently across time (for example, functional brain development [29], [30]). To this end, in this work, we propose a new TXAI system, called TXAI, characterized with time-conditioned distribution for analyzing a time-variant real-life process across time.

The proposed TXAI system can delineate the trajectories of a dynamic, real-life process across time. In addition, a comparison with state-of-the-art AI systems, with varying levels of explainability, manifested that the proposed TXAI performed better than most of the compared AI systems (for, e.g., for mean f-score TXAI performs better than LSTM by 18.19%, DT by 6.81%, HMM by 4.90%, GT2-based XAI system by 8.58% on test datasets) except TCN, which is a much more complex, and a black-box method. XAI systems based on standard FLS (e.g., T1, IT2, or GT2) are unable to integrate information relative to the time dimension. More specifically, the TXAI system credit the membership value of a fuzzy concept given the fuzzy concept using conditional distribution. The conditional distribution is then utilized to investigate the evolution of the process across different time intervals. In this way, the TXAI is able to predict the likelihood of observing prototypical rules of the process across different time intervals. For future works, the proposed TXAI system can have profound implications to contribute to our understanding of temporal real-life processes, for instance, human-centred systems and life sciences. Furthermore, for these future life science studies, we would also endeavor that TXAI entails all ethical concerns accounted for a more fair, and complete TXAI analysis.

ACKNOWLEDGMENT

The authors would like to thank Oracle for Research staff in Europe, Richard Pitts, and Michael Riley, for their support via Oracle Cloud credits and technical assistance relating to this research.

REFERENCES

[1] M. S. A. Lee, L. Floridi, and A. Denev, Innovating With Confidence: Embedding AI Governance and Fairness in a Financial Services Risk Management Framework. Berlin, Germany: Springer, 2021, ch. 9, pp. 353–371.
[2] K. J. Yaxley, K. F. Joiner, and H. Abbass, “Drone approach parameters leading to lower stress sheep flocking and movement: Sky shepherding,” Sci. Rep., vol. 11, pp. 1–9, 2021.
[3] R. Li, D. Auer, C. Wagner, and X. Chen, “A generic ensemble based deep convolutional neural network for semi-supervised medical image segmentation,” in Proc. IEEE 17th Int. Symp. Biomed. Imag., 2020, pp. 1168–1172.
[4] M. Kiani, J. Andreu-Perez, H. Hragas, E. I. Papageorgiou, M. Prasad, and C. T. Lin, “Effective brain connectivity for fNIRS with fuzzy cognitive maps in neuroergonomics,” IEEE Trans. Cogn. Develop. Syst., vol. 14, no. 1, pp. 50–63, Mar. 2022.
[5] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, “Brain intelligence: Go beyond artificial intelligence,” Mobile Netw. Appl., vol. 23, pp. 368–375, 2018.
[6] J. Andreu-Perez et al., “A generic deep learning based cough analysis system from clinically validated samples for point-of-care covid-19 test and severity levels,” IEEE Trans. Serv. Comput., vol. 15, no. 3, pp. 1220–1232, May/June 2022.
[7] H. Hragas, “Toward human-understandable, explainable AI,” Computer, vol. 51, no. 9, pp. 28–36, 2018.
[8] G. Alicioglu and B. Sun, “A survey of visual analytics for explainable artificial intelligence methods,” Comput. Graph., vol. 102, no. 9, pp. 502–520, 2022.
[9] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 1, pp. 28–44, Jan. 1973.
[10] L. A. Zadeh, “Fuzzy sets,” Inf. Control, vol. 8, no. 3, pp. 338–353, 1965.
[11] L. M. Candanedo and V. Feldstein, “Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models,” Energy Buildings, vol. 112, pp. 28–39, 2016.
[12] J. M. Garibaldi, M. Jaroszewski, and S. Musiakowski, “Nonstationary fuzzy sets,” IEEE Trans. Fuzzy Syst., vol. 16, no. 4, pp. 1072–1086, Aug. 2008.
[13] A. V. Kostikova, P. V. Tereliansky, A. V. Shuvaev, V. N. Parakhina, and P. N. Timoshenko, “Expert fuzzy modeling of dynamic properties of complex systems,” ARPN J. Eng. Appl. Sci., vol. 11, no. 17, pp. 10222–10230, 2016.
[14] H. Maeda, S. Asakoa, and S. Murakami, “Dynamical fuzzy reasoning and its application to system modeling,” Fuzzy Sets Syst., vol. 80, no. 1, pp. 101–109, 1996.
[15] D. Kozen and M. Timme, “Indefinite summation and the Kronecker delta,” 2007. [Online]. Available: https://hdl.handle.net/1813/8352
[16] C. Wagner and H. Hragas, “Toward general type-2 fuzzy logic systems based on zSlices,” IEEE Trans. Fuzzy Syst., vol. 18, no. 4, pp. 637–660, Aug. 2010.
[17] Q. Liang and J. Mendel, “Interval type-2 fuzzy logic systems: Theory and design,” IEEE Trans. Fuzzy Syst., vol. 8, no. 5, pp. 535–550, Oct. 2000.
[18] C. Chen, D. Wu, J. M. Garibaldi, R. I. John, J. Twycross, and J. M. Mendel, “A comprehensive study of the efficiency of type-reduction algorithms,” IEEE Trans. Fuzzy Syst., vol. 29, no. 6, pp. 1556–1566, Jun. 2021.
[19] M. Nie and W. W. Tan, “Towards an efficient type-reduction method for interval type-2 fuzzy logic systems,” in Proc. IEEE Int. Conf. Fuzzy Syst. (IEEE World Congr. Comput. Intell.), 2008, pp. 1425–1432.
[20] S. Mirjalili, “Genetic algorithm,” in Evolutionary Algorithms and Neural Networks, Berlin, Germany: Springer, 2019, pp. 41–55.
[21] D. P. Filev and I. Kolmanovsky, “Generalized Markov models for real-time modeling of continuous systems,” IEEE Trans. Fuzzy Syst., vol. 22, no. 4, pp. 983–998, Aug. 2014.
[22] F. Herrera, “Genetic fuzzy systems: Taxonomy, current research trends and prospects,” Evol. Intel., vol. 1, no. 1, pp. 27–46, 2008.
[23] A. Shershinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” Physica D: Nonlinear Phenom., vol. 404, 2020, Art. no. 132306.
[24] S. Fine, Y. Singer, and N. Tishby, “The hierarchical hidden Markov model: Analysis and applications,” Mach. Learn., vol. 32, pp. 41–62, 1998.
[25] Y. Ben-Haim and E. Tom-Tov, “A streaming parallel decision tree algorithm,” J. Mach. Learn. Res., vol. 11, no. 2, pp. 849–872, 2010.
[26] P. Remy, “Temporal convolutional networks for keras,” 2020. [Online]. Available: https://github.com/philipperemy/keras-tcn
[27] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.
[28] H. R. Rocha, I. H. Honorato, R. Fiorotti, W. C. Celeste, L. J. Silvestre, and J. A. Silva, “An artificial intelligence based scheduling algorithm for demand-side energy management in smart homes,” Appl. Energy, vol. 282, 2021, Art. no. 116145.
[29] M. Kiani, J. Andreu-Perez, H. Hragas, S. Rigato, and M. L. Filippetti, “Towards understanding human functional brain development with explainable artificial intelligence: Challenges and perspectives,” IEEE Comput. Intell. Mag., vol. 17, no. 1, pp. 16–33, Feb. 2022.
[30] J. Andreu-Perez, L. L. Emberson, M. Kiani, M. L. Filippetti, H. Hragas, and S. Rigato, “ Explainable artificial intelligence based analysis for interpreting infant fNIRS data in developmental cognitive neuroscience,” Commun. Biol., vol. 4, no. 2021, Art. no. 1077.