Spatiotemporal Relationships Extraction from ISH Gene Expression Data

Modeling of Fuzzy and Possibilistic Association Rules

Noureddine Mekroud
Department of Computer Science, University Setif 1 - Ferhat Abbas, Setif, Algeria
mekroud_n@univ-setif.dz (corresponding author)

Abdelouahab Moussaoui
Department of Computer Science, University Setif 1 - Ferhat Abbas, Setif, Algeria
abdelouahab.moussaoui@univ-setif.dz

Received: 23 April 2025 | Revised: 7 June 2025 | Accepted: 18 June 2025

 DOI: https://doi.org/10.48084/etasr.11614

ABSTRACT

Imperfection is a common feature in almost all real-world data, although it usually hides crucial knowledge of major interest. Using theories of uncertainty to model gene expression (gen-exp) areas in the embryo can help extract hidden relationships between genes while taking into account possible imprecision in the boundaries of gen-exp zones and their nuanced strength. This study proposes fuzzy and possibilistic modeling of spatiotemporal data from In-Situ Hybridization (ISH) sequences of images representing gen-exp areas in different embryonic developmental phases of the model species Edinburgh Mouse. Following a series of preprocessing steps on these images to improve feature extraction, an adaptation of the Apriori algorithm to the fuzzy and possibilistic logic is proposed for mining two types of Association Rules (AR) that represent the spatial correlations between gen-exp areas in the embryo, and the temporal correlations as well as the relationships between genes that co-express in these ISH image sequences. Finally, an analysis of extracted spatiotemporal fuzzy and possibilistic AR is provided to decide which modeling is the most suitable. Biological interpretation of the results obtained confirms their adequacy with the domain principles. The extracted knowledge can help biologists better understand the interactions between genes, discover the coexpressed sets of genes that have the same functional role, and model normal gen-exp for detecting abnormal gen-exp that can cause genetic diseases.

Keywords-association rules; spatiotemporal data; ISH images; gene expression; fuzzy and possibilistic logic

I. INTRODUCTION

The phenomenal increase in the volume of spatiotemporal data produced in different research domains has generated various problems in modeling and interpreting them, concealing a large amount of knowledge [1]. Bioinformatics uses computational and soft computing methods [2] to process vital heterogeneous biological data, which are expanding extensively but are often imprecise and/or incomplete, and their interpretation is still a challenge. Gene expression (gen-exp) data integrate crucial knowledge about the stages of embryonic development of any model species, but mainly, this knowledge remains undiscovered and, therefore, unexploited by biologists. Among the powerful forms of representation of gen-exp are ISH image sequences, which indicate in situ areas where each gene is expressed in the embryonic growth phases of some typical species [3-5]. These images represent a spatiotemporal biological dataset, and their study is essential for the pre-detection of diseases due to genetic causes during the preliminary developmental phases of living species.

Association Rules (AR) [6] are one of the most powerful algorithms in machine learning, used to extract hidden relationships between a set of attributes initially in binary format. However, in the real world, humans generally deal with ambiguous knowledge about any situation, either because they have doubts about its veracity (uncertainty) or because they find it difficult to state clearly (imprecision) [7]. Therefore, binary representation of real-world data generally causes a great loss of precision in their quantification and semantics. Thus, other AR versions have been proposed to make them closer to human reasoning and the reality of the analyzed data by modeling sequential [8] and quantitative [9] data, or supporting nuanced data through fuzzy-based modeling [10], possibilistic-based modeling [11], and evidential-based modeling [12]. These adapted AR help extract more relevant knowledge. A good summary of AR was presented in [13].

AR mining machine learning algorithms are used to analyze gen-exp data, especially for the selection of the most interesting AR based on measurements of objective interest (purely computational interest) [14], or subjective interest (biologically

interpretable AR) [15, 16]. Three main technologies have been presented to sequence gen-exp data:

- DNA microarray data: Most works focus on a selection of interesting features for efficient gene classification [17] or search for adequate distances and similarity measures to cluster and discover hidden patterns [18].
- RNAseq (RNA sequencing): Works focusing on feature selection for classification [19] or clustering [20].
- ISH images (DNA or RNA sequences of high-resolution images) are studied for the detection of genes that co-express [21] or clusters to extract hidden patterns [22].

Regarding ISH gen-exp data, in [14-16], only binary representation was used for modeling, which can severely influence the quality of the extracted knowledge because the studied data were considered precise and perfect. However, in reality, these data can hide forms of uncertainty (it is not strictly certain that a gene is always expressed in exactly such a zone of pixels in such a phase at such a level of expression) and imprecision (the boundaries of gen-exp zones are not surely strictly defined). Therefore, ignoring a possible uncertainty or imprecision in these data can decrease the relevance of the results of the proposed approaches.

Little effort has been put into the extraction of AR from gen-exp data [10, 23], and no fuzzy AR modeling has been proposed to analyze ISH data, which are uncertain and imprecise by nature (see Figure 1). Also, to our knowledge, no published work has studied this type of gen-exp data by formulating them via a possibilistic logic, which is a broader vision of fuzzy logic, to analyze them using any computational method, including AR. In these perspectives, this work presents an approach adjusted to the biological reality of ISH gen-exp images resulting from the monitoring of the development stages of the Edinburgh mouse model species [3]. The focus is on adapting the Apriori algorithm to theories of uncertainty for modeling the analyzed data to extract several types of useful AR, according to the two proposed models (fuzzy and possibilistic). In addition, the nuanced level of gen-exp (strong, moderate, and not detected) is taken into account to extract the most interesting features that represent the reality of these biological spatiotemporal data.

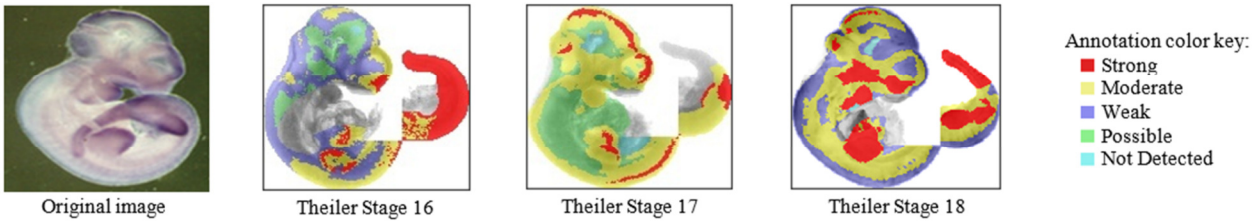


Fig. 1. Example of gen-exp images of gene named "ATF4" mapped to the original image of a standard Edinburgh Mouse embryo (image on the left) during development phases TS16, TS17, TS18 (Reference: EMAGE 3052 in the online database: www.emouseatlas.org). The zones of gen-exp are colored according to the level of their expression (see legend: Annotation color key)

After a sequence of preprocessing operations and feature extraction to reduce the dimensionality of studied images (and the algorithmic complexity) while keeping the maximum of their variability, the aim is to extract relationships between gen-exp by detecting sets of genes that co-express, regions that undergo a synchronized gen-exp, and temporal relationships between genes that express in different stages of embryonic development. Extracted AR helps to discover several biological knowledge hidden in these ISH images.

The proposed adaptation of the Apriori algorithm (initially designed for binary data) is based on the redefinition of notions of cardinality and union between itemsets. Also, a method based on triangular norms is used for calculating the support of itemsets and the confidence of extracted AR, to make them suitable for fuzzy and possibilistic logics.

Finally, an analysis and comparison between the quality of extracted rules from the two fuzzy and possibilistic models is proposed to decide which modeling is the most adequate. Also, a functional analysis of the genes forming the itemsets of extracted AR confirms that a large part of the detected co-expressions is in accordance with biological principles, which affirms that these genes really collaborate during the formation of different biological systems.

II. APPLICATION DATA

The images of gen-exp represent the expression strength (level) of genes during different development stages of any model species embryo. The most studied typical species are:

- The Edinburgh Mouse (EMAGE database) [3].
- The Zebrafish (ZebraFish Information Network, ZFIN) [4].
- The Fruit Fly Drosophila (Drosophila FlyBase) [5].

The studied dataset EMAGE (Edinburgh Mouse Atlas of Gene Expression) is a collection of high-quality ISH images that aims to capture: spatial location (areas) of gen-exp levels in the embryo morphology (tissues, organs, or whole body), and temporal phases when genes are expressed, named Theiler Stages (days).).). Thus, EMAGE integrates various hidden spatiotemporal gen-exp patterns.

For technical and economic reasons, the mouse has become by far the animal most used by researchers to study the impact of new drugs and genetic therapies by applying them over several generations of these animal models, given the rapid growth of its embryo (19 to 23 days). In addition, the mouse genome is entirely sequenced and is very close to that of man [3].

III. THEORETICAL BACKGROUND

A. Theories of Uncertainty

In the real world, we frequently have to process ambiguous data of uncertain and/or conflicting nature, either because there is doubt about their veracity (uncertainty), or we encounter difficulties in stating them clearly (imprecision), or they are simply contradictory (inconsistency) [7].

In binary logic (the most basic), the degree of veracity of a proposition can only take one of two values {1, 0} (true or false), which is not suitable for modeling (and processing) the majority of real-world data, which is far from being binary. In [24], a new theoretical framework was developed to model uncertainty, the theory of Fuzzy Sub-Sets (FSS), which allows manipulation of nuanced veracity between completely true and completely false (partial truth in the interval [0, 1]) for no hard separation of overlapped classes. The FSS logic has been extended to possibilistic logic [25], which was released from the wedging constraint FSS (as well as binary) logics, which requires that the sum of degrees of membership of an object to the set of classes must always be equal to 1. In addition, evidential logic (Dempster-Shafer) based on belief functions [26] allows modeling conflicts between multiple data sources. The following subsections focus on the theories of FSS and the possibilistic logic used in this work.

B. Basic Concepts of the FSS Theory

The FSS theory accommodates the concept of partial truth (or gradual veracity). For FFS modeling, we must define [24]:

- Universe of Discourse (UoD) includes the terms used in a studied field, e.g., temperature is expressed using hot, cold, lukewarm. Each of these terms will form an FSS and is represented by a membership function.
- Membership Function (MsF)*:* An FSS A in a UoD X is characterized by its MsF $\mu(A)[x]$, which associates each element x (of X) a real value in the interval [0, 1], and measures the degree to which x belongs to the set A.

C. Basic Concepts of the Possibility Theory

Introduced in [25] as an extension of FSS theory and then developed by [27], the possibility theory is based on:

- Measure of possibility Π*:* For a reference set X, Π assigns to each subset $A \in X$ a real number in [0,1], evaluating at which point an event A_i is possible, such as: $\Pi(\emptyset) = 0$, $\Pi(X) = 1$ and $\forall A_i \in X,\ \Pi(\cup\ (A_i)) = Max(\Pi(A_i))$. This function expresses the possibility of realizing one of the events A_i, taken indifferently, is the same as realizing the event as much as possible between them.
- Measure of necessity N*:* For a reference set X, N indicates the degree to which the realization of an event A_i is certain. N is the dual magnitude of Π, which assigns to all A a real in [0,1], and further verifies: $N(\emptyset) = 0$, $N(X) = 1$, and $\forall A_i \in X,\ N(\cap\ (A_i)) = Min(N(A_i))$. N can be obtained from Π by: $\forall\ Y_i \in X,\ N(\ Y_i) = 1 - \Pi(\ {Y_i}^c)$. Thus, knowing that Y_i^C is the opposite event of Y_i, the more Y_i is necessary, the less ${Y_i}^C$ is possible, which makes it possible to evaluate the certainty of the realization of Y_i.

D. Association Rule Mining

Initially developed for the analysis of sales datasets to identify sets of items purchased most frequently together, the basic elements of the AR algorithm are [28]:

- Item, Itemset: In transaction databases, an item is a pair (attribute, value) with values belonging to a defined domain. A non-empty set of items constitutes an itemset.
- Support of itemset: percentage of transactions (rows) in the database supporting (containing) simultaneously all its items:

$$Supp(AB) = P(AB) = \frac{nb(AB)}{nb(T)} \quad (1)$$

  where $nb(AB)$ is the number of rows where A and B appear together, and $nb(T)$ is the total number of transactions T in the database. In a frequent itemset, its $Supp \geq min\_threshold$, noted $MinSupp$.
- An AR has the form A→ B [*Supp%*, *Conf%*]. A, B are attributes (items or itemsets, A∩B=Ø), which means that if A is present in a transaction, B is likely present too. The itemset A is called the antecedent (or condition), and the itemset B is called the consequent (or result).
- Confidence is the proportion of transactions that perform B among those that perform A:

$$Conf(A \rightarrow B) = \frac{P(AB)}{P(A)} = \frac{nb(AB)}{nb(A)} \quad (2)$$

- Lift is a measure of the importance of AR, defined by:

$$Lift(A \rightarrow B) = \frac{Conf(A \Rightarrow B)}{Supp(B)} = \frac{P(AB)}{P(A)P(B)} = \frac{nb(AB)nb(T)}{nb(A)nb(B)} \quad (3)$$

- The Apriori algorithm is a reference for extracting frequent itemsets for binary AR [28], using an iterative approach known as research by level, where the frequent k-itemsets (in level k) forming the set $L(k)$ will be used to create the following candidate $C(k+1)$ itemsets, and keeping only the frequent $(k+1)$-itemsets in $L(k+1)$ whose $Supp \geq MinSupp$.

Finally, from the frequent itemsets discovered, another process will generate a valid AR having *Conf* ≥ *MinConf* threshold.

E. Adaptation of AR Mining for Fuzzy Logic

Initially, Apriori was unsuitable to deal with nuanced data. In [29], AR are defined as integrating concepts of FSS theory. Fuzzy AR has the form "If X is A (antecedent), then Y is B (consequent)". This AR is noted as $(X, A) \rightarrow (Y, B)$, such as $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$ are two disjoint itemsets, and $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_m\}$ are FSS associated with the elements of X and Y. Here are some useful definitions:

- Fuzzy-item is a pair (item, FSS), e.g., (*chocolate*, *much*), where *much* is an FSS defined by its MsF.
- Fuzzy-itemset is a set of fuzzy-items, e.g., [(*chocolate*, *much*); (*milk*, *little*)], where *much* and *little* are two FSS.

- Degree of Membership (DoM) μ: each fuzzy-item (x_i, a_j) is represented in a transaction t by its DoM denoted $(\mu(a_j)t[x_i]) \in [0,1]$. That means that the DoM of item x_i to a_j at the row t is $\mu(a_j)$. A fuzzy-item or fuzzy-itemset is said to be frequent if its $DoM \geq \Omega$ (min-threshold set by an expert). The calculation of the DoM of an itemset from the DoMs of its items is described below.
- $Fuzzy - AR(X, A) \rightarrow (Y, B)$ is valid if a sufficient number of rows in T support (contain) the pairs $(X, A) \wedge (Y, B)$.

Therefore, adaptation of AR mining for fuzzy logic requires a redefinition of basic operations in classical set theory to deal with overlapped sets, especially the union of items (to create a fuzzy-itemset and deduce its DoM) and the cardinality (to count the frequency of appearance of itemsets in all transactions). Support of fuzzy itemsets, as well as confidence of the extracted fuzzy-AR, will then be deduced from the redefinitions of these two basic operations [30]:

- Union of fuzzy-items: to calculate the DoM of fuzzy-itemsets, the notion of triangular norms $\mathcal{T}$ is used, noted $t - norm(T)$ as a generalization of the intersection operator, and $t - conorm(\perp)$ as a generalization of the union operator. This study uses one of the most used t-conorms to calculate the DoM of n -items, the max approach, defined as:

$$\text{Max}_{i=1..n}(\mu(a_j)t[x_i]) = \perp_{j=1..m}(\mu(a_j)) = \text{Max}_{j=1..m}(\mu(a_j)) \quad (4)$$

- Support of fuzzy-itemset: Denoted $Fuzzy\_supp(X, A)$, it is the percentage of transactions supporting (containing) the $fuzzy - itemset(X, A)$ out of the total number of transactions. The calculation of the fuzzy-itemset support is based on a redefinition of the cardinality of fuzzy-itemsets, which is their frequency of appearance in the rows of the transaction table. The cardinality of $fuzzy - itemset(X, A)$ is closely related to its DoM $\mu(a_j)$ calculated using several approaches [30], and the most suitable for this modeling (based on thresholds) is sum threshold counting, by adding only DoMs (of each row containing this itemset) that exceed a threshold w, so:

$$Fuzzy\_supp(X, A) = \frac{\sum_{t \in T} \mathcal{T}(\mu(a_j)t[x_i])}{nb(T)} \quad (5)$$

where t having $Max_{i=1..n}(\mu(a_j)t[x_i]) > w$, with $\mathcal{T}$ being a triangular norm to calculate the DoM of fuzzy-itemsets.

- The confidence of the Fuzzy-AR is calculated by:

$$Fuzzy_{conf((X,A)\rightarrow(Y,B))} = \frac{Fuzzy\_supp(X \cup Y, A \cup B)}{Fuzzy\_supp(X, A)} \quad (6)$$

IV. PROPOSED APPROACH

The proposed approach consists of extracting fuzzy and possibilistic AR from sequences of ISH images of gen- exp. It is an improvement over previous works that tried to study the same type of data through binary logic [14-16], which is inadequate to their reality and causes a great loss of hidden biological knowledge. After a series of preprocessing operations to vectorize ISH images for extracting local features, fuzzy and possibilistic modeling is used to represent more reliably the reality of the studied data and extract relevant and biologically plausible knowledge.

A. The Proposed FSS Model

Each of the three expression levels, strong (red color), moderate (yellow), and not-detected (gray color), is represented by an FSS and linked to an MsF (Figure 2), which affects a gradual DoM to each FSS. This means a formulation of DOM according to the gen-exp strength of 6×6 pixels of each case in the ISH image to extract features.

The studied images were vectorized according to the FSS modeling by converting the images to 256 grayscale images (to define the UoD). For the results from experimental tests of pixel gray levels in each of the three gen-exp strengths, the intervals of kernels of the three MsF following their grey levels are: strong [0-76], not detected [128-128], and moderate [225-255]. For example, if the grey level of a pixel is equal to 100, it is considered to belong to a strong FSS with a DoM of 0.5 and a not-detected FSS with a DoM of 0.5, knowing that the sum of all DoMs must be equal to 1.

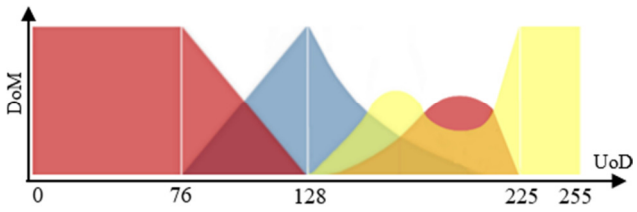


Fig. 2. MsF defined for the proposed FSS modeling: FSS_{Strong} in red, $FSS_{Moderate}$ in yellow, $FSS_{Not\text{-}Detected}$ in gray.

According to this modeling, any case of 6×6 pixels in ISH images will be an attribute (a column in the transaction table), and each image will be a row. Also, each cell in this table will contain three values representing the DoM of a case to each of the three FSS, following the MsF shown in Figure 2.

B. The Proposed Possibilistic Model

To define a function to calculate the DoM for each FSS according to possibilistic reasoning (which does not require that the sum of the three DoM must be equal to 1), some statistical estimators were used, such as Maximum Likehood (MLE) and Maximum A Posteriori (MAP). MAP is an estimator of a certain number of unknown parameters θ, such as the density of probability P related to a given sample X. In this possibilistic modeling, the basic formula of MAP is [31]:

$$\theta_{MAP}(X) = arg_\theta \, max \, \rho(X/\theta)\rho(\theta) \quad (7)$$

Classification by the MAP method seeks the most probable class C_i ($C_i = \theta$, an FSS here) given X (a vector of X_j), therefore, the class which maximizes $P(X/C_i)$. MAP is a probabilistic method based on Bayes' rule:

$$\rho(C_i/X) = \rho(X/C_i) \times \rho(C_i)/\rho(X) \quad (8)$$

where $P(C_i/X)$ is the probability of class C_i having the X case, $P(X/C_i)$ is the probability of case X (of each of its 4 zones X_j) to belong to class C_i (the MAP method will choose the max), $P(C_i)$ is the probability of belonging to class C_i (independent), and $P(X)$ is the probability of being case X.

For each zone X_j, $P(X_j) = 1/4$ (a constant). Therefore, $P(C_i)/P(X)$ is constant for all X_j. Thus, $P(C_i/X)$ depends on $P(X/C_i)$ only, and then:

$$P(C_i/X) = MAP\left(P(X_j/C_i)\right) \quad (9)$$

The proposed possibilistic vectorization is as follows: each case X of 6×6 pixels is divided into four zones X_j of 3×3 pixels. Then, according to the MAP method, for each zone X_j, the probability of belonging to each of the three FSS is calculated, with their maximum being the DoM of this case X in the image (cell in the table) to an FSS, respectively.

C. Flowchart of the Proposed Approaches

When the DoM to an FSS of a zone (and thus of a case) is equal to 1, this case is said to be homogeneous since it contains a zone that has undergone clear gen-exp in one of the three levels of expression. Figure 3 illustrates the preprocessing operations performed, the uncertainty modeling proposed to extract the two transaction tables, and deduces the most relevant AR (Fuzzy-AR or Possibilistic-AR).

D. Contribution

This approach presents a modeling that is adequate to the reality of the studied data to ensure the extraction of hidden knowledge with clear biological significance, unlike previous works [14-16] that treated gen-exp in binary logic by reducing the forms of expression to only detected or not detected and neglecting all other possible forms. This approach studies the three most significant forms of gen-exp (strong, moderate, and not detected) to retain the most relevant knowledge. The other forms of expressions (weak and possible) only increase the algorithmic complexity of the Apriori algorithm without carrying reliable information. In addition, although the EMAGE database provides high-quality images, the risk of imprecision during the spatial annotation can not be ignored in presuming to have detected exactly the borders of colored areas (levels of gen-exp). Thus, to model more precisely the reality (far from being binary and precise) of the biological data:

- The proposed FSS model provides a nuanced assignment of each 6×6 pixels in an ISH image to each of the three forms of expression through an MsF for each of the three FSS.
- The model is extended to possibilistic logic for solving the problem of the blurring effect often caused by the nuanced assignment of each case through FSS modeling. This effect erases the boundaries between areas of gen-exp, setting aside all forms of distribution (concatenated or dispersed) of colored pixels in a case. Through the MAP method, the boundaries of areas of expression are preserved, and the areas that have undergone a clear gen-exp (concatenated pixels of the same color) are favored. In addition, the computation of the DoM via MAP gives a remedy for noisy pixels by favoring the most homogeneous adjacent pixels.

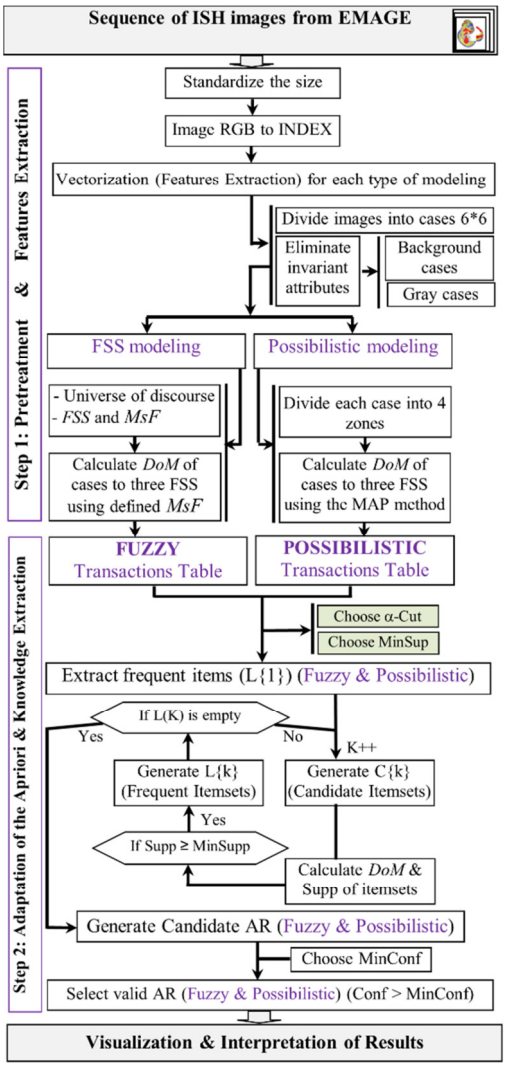


Fig. 3. Flowchart of the two proposed approaches.

These two approaches help in discovering three different AR forms, according to two points of view (primal and dual):

- Extract AR between areas (spatial relationships): When adjacent pixels (organ) undergo a gen-exp, other organs in biological relation should also undergo a gen-exp according to that level of expression.
- Extract AR between genes: When a gene is expressed, other genes must be expressed as well.
- From AR between genes, and since a nomination combines the name of each gene with the TS it expresses, a new form of temporal knowledge is extracted, helping to understand the relationships between phases of gen-exp (embryo development phases).

Despite the combinatorial explosion of possible subsets of genes, the obtained results show that with MCT=90%, and the rate of gene relationships (in the extracted AR) recognized by the platform String-DB.org [32] is 37.74% for the Fuzzy-AR and 42.56% for the Possibilistic-AR. Finally, 62.94% of Fuzzy-

AR present temporal relationships between the same TS phases, and 37.06% are between different TS phases. For Possibilistic-AR, 77.81% are between the same TS phases, and 22.19% are between different TS phases.

V. IMPLEMENTATION AND RESULTS ANALYSIS

A. Preprocessing and Dimensionality Reduction

The preprocessing operations reduced the dimensionality of the studied data while retaining the maximum variability of their content. For feature extraction, the proposed modeling helped to reliably translate the reality of these data through their vectorization toward two transaction tables by:

- Compress each image of 150×150 pixels to an image of 25×25 nuanced cases of 6×6 pixels (from 1502 to 252 variables, dimensionality reduction of 97.3%).
- Remove invariant attributes that do not carry knowledge, representing white cases (background) and cases always gray (have not undergone any gen-exp in all 1694 images). In addition, eliminate attributes (columns) with weak expression (having a DoM in all their cases under the α-cut of the MsF).

The Apriori algorithm has a hidden mechanism to reduce its algorithmic complexity: frequent itemsets are generated by frequent itemsets as well (see Figure 3).

B. Adjustment of Parameters and Generation of Transaction Tables

A vast set of empirical tests was carried out to choose adequate values for the following hyperparameters:

- α-cut: to retain the most relevant attributes (items), an α-cut $\Omega = 0.3$ was chosen for the MsF as minimum DoM.
- Min-Supp: to select the frequent itemsets, Min-Sup=0.3 for Fuzzy-AR and 0.43 for Poss-AR were chosen to avoid a combinatorial explosion.
- Min-Conf: to extract the most relevant AR, it was set to 0.9.

After the preprocessing process described in Figure 4, each image in the EMAGE database, representing the expression of a gene, becomes a row in the resulting table. The two transaction tables resulting from the proposed preprocessing process have the following structure:

- Each line (transaction i) represents the vectorization of a gene expression image i, so there are 1694 lines.
- Each column (attribute j) represents a feature of 6×6 pixels; therefore, there are 625 columns.
- Each case (i, j) is a cell in the table, composed of 3 DoM (see the example of a cell at the end of Figure 4).

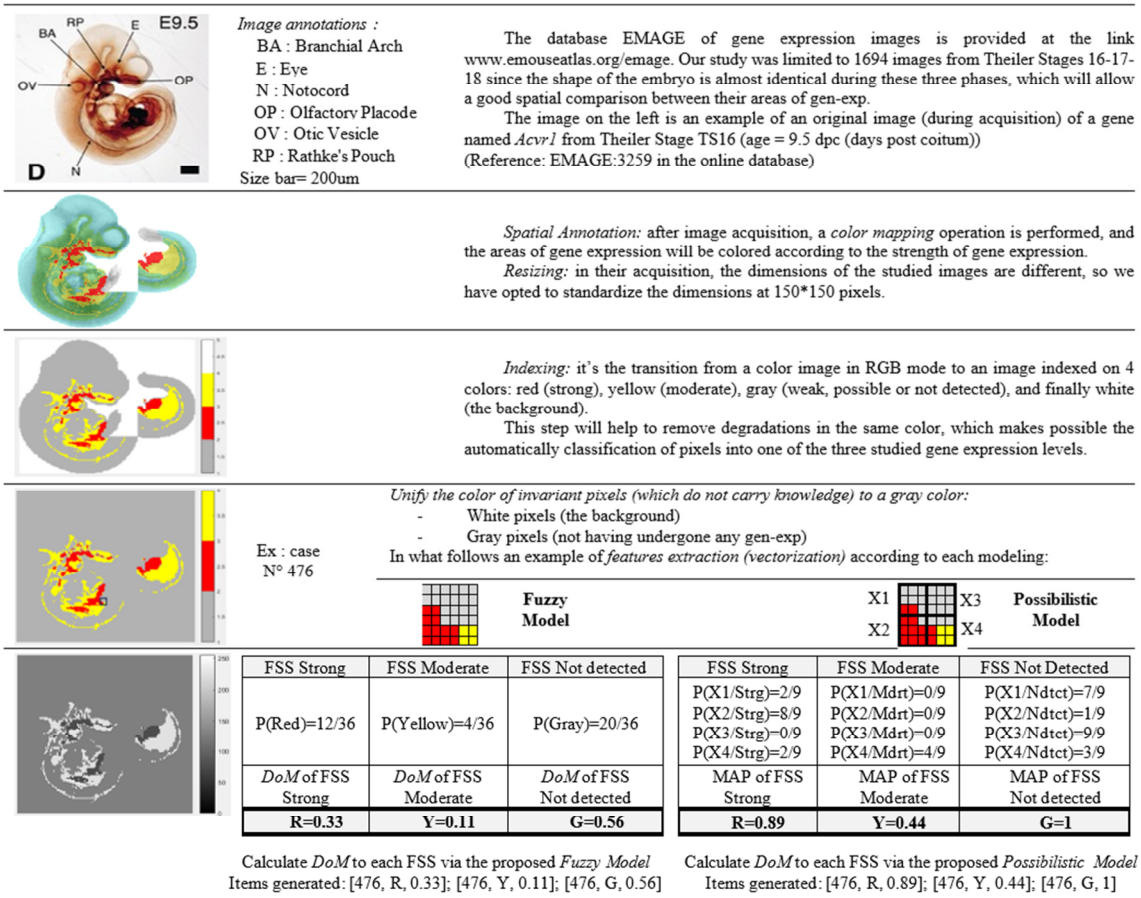
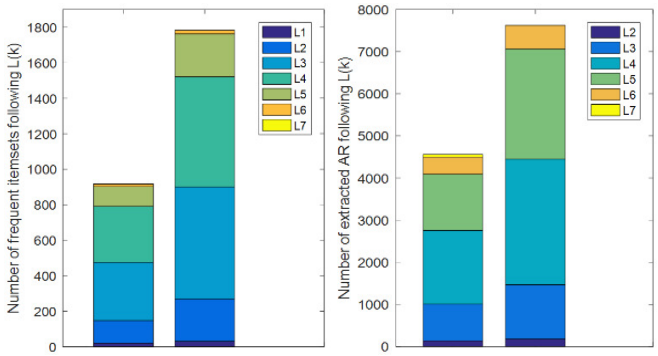


Fig. 4. Sequence of preprocessing operations for extracting fuzzy and possibilistic transaction tables from the studied ISH images.

C. Analysis and Discussion of Spatial AR Extracted in Primal

Extracted AR represent spatial relationships (AR between cases), temporal relationships between the developmental phases of the embryo (Theiler Stages (TS)), and relationships between co-expressing genes. The histograms below visualize the generated frequent itemsets, following their $L(k)$, and extracted AR of the proposed two approaches.



(a) (b)

Fig. 5. Frequent fuzzy and possibilistic itemsets and extracted AR in primal: (a) Fuzzy MST = 0.3, Possibilistic MST = 0.37; (b) Fuzzy MST = 0.3, Possibilistic MST = 0.43.

From these results, it can be observed that fuzzy modeling generates a limited number of frequent itemsets, and so few fuzzy AR will be generated. However, the possibilistic modeling generated more frequent itemsets, and so more AR. These can be explained by the MAP principle, which chooses the max of DoM to each FSS between the four subcases, which makes evident that more possibilistic AR will be generated. Also, some fuzzy or possibilistic AR have *Conf* and *Lift* >1, which is also explained by the principles of MAP and max *t*-conorm approaches used in this possibilistic model, which always choose the max of DoM in each case.

TABLE I. EXAMPLES OF STRONG FUZZY AND POSSIBILISTIC AR IN PRIMAL

	Antecedent	Consequent	Supp	Conf	Lift	Euclidean distance
Fuzzy AR	233^510	263	0.30	0.96	2.16	11.18
	484	263^485^510	0.30	0.93	2.73	10.44
Possibilistic AR	237^510	262^263	0. 37	0.98	1.85	11.18
	233^258^485	234	0.38	0.99	2.10	10.20

To focus on AR that are the most carriers of novelty and relevant knowledge, obvious AR were filtered out, resulting from:

- Adjacent cases: Since a gen-exp area can affect many adjacent cases in the embryo, this causes the extraction of AR between adjacent cases with a small Euclidean distance.
- Overlapped AR: AR having common sub-itemsets both in their antecedent and consequent parts.
- Whole red images: Having undergone strong gen-exp in large areas (caused by dominant genes), these cases will be extracted (imposed) in a large number of biased AR.

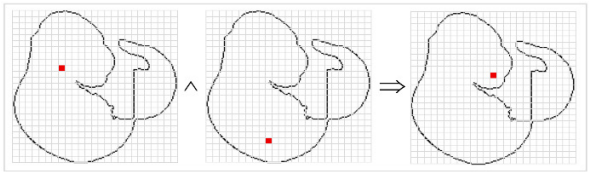


Fig. 6. Example of spatially extracted AR (between cases - organs).

D. Interest of Duality in the Proposed Approach

Following the principle of duality, two new tables can be generated, which are the transposed of the two transaction tables, where a column (attribute) will represent a gene (image of gen-exp) and the row will represent the cases of the studied images. This duality allows the extraction of crucial AR representing the relationships between genes. Also, given the combined nomination [TS-Gene] of gen-exp images, these gene-gene AR will clarify the temporal sequencing of gen-exp during the embryo developmental phases (Table II, Figure 7). Figure 7 represents an example of AR between genes, through the TS16 and TS17 developmental phases, extracted through the principle of duality.

TABLE II. EXAMPLE OF FUZZY AND POSSIBILISTIC AR IN DUAL

	Antecedent	Consequent	Supp	Conf	Lift
Fuzzy AR	TS16_Zfp410_Strong	TS17_Zfp623_Strong	0.36	1.10	3.33
	TS17_Skp1a_Moderate	TS16_Zfp410_Strong TS17_Zfp623_Strong	0.30	0.99	2.74
Possibilistic AR	TS17_Zfp623_Strong	TS16_Zfp410_Strong	0.48	0.97	1.92
	TS16_Nfatc2_Moderate TS16_Zfp410_Strong	TS17_Zfp623_Strong	0.43	0.90	1.82

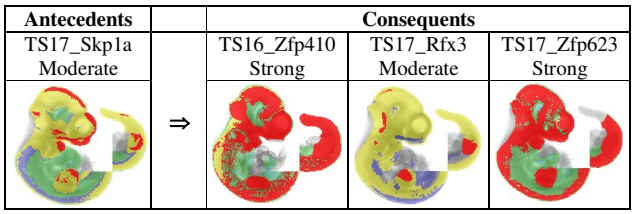
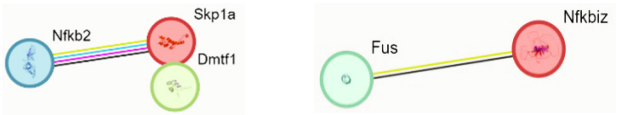


Fig. 7. Example of Fuzzy-AR in Dual (MST=0.28, MCT=0.93, Lift=2.89).

E. Biological Interpretation of the Obtained Results in Dual

Figure 8 represents two graphs of dependency, recognized by the string-db.org [32] platform, simulating the strength of co-expression between genes of frequent itemsets or extracted AR. String-db.org is a biological database for collecting gene-gene or protein-protein interactions recognized and predicted by researchers. A significant part of the extracted AR is recognized in string-db.org, which proves the relevance of these results. However, if some extracted AR is not recognized by this platform, it does not affect the credibility of the results (supported by solid metrics: Supp, Conf, Lift), but it only

means that string-db does not yet predict these genes' co-expressions. Such AR can carry biological novelties not yet known to biologists, which is the most important asset of this work.



(a) (b)

Fig. 8. Examples of itemsets and AR recognized by string-db.org [32] (Strong recognition means more arcs between nodes - genes): (a) Example of fuzzy extracted AR recognized partially in string-db.org, (b) Example possibilistic frequent itemset recognized entirely in string-db.org.

VI. CONCLUSIONS

The objective of this work was to model the uncertain aspect in gen-exp data for extracting spatiotemporal fuzzy and possibilistic AR to provide a formal description of hidden relationships between genes in sequences of ISH images of the model species Edinburgh Mouse, which is a rich source of crucial knowledge about embryonic development. Initially, a series of preprocessing operations was applied to extract the most relevant features in these images, according to several criteria: level of gen-exp, variability, and α-cut of DoM in fuzzy and possibilistic modeling. Thus, fuzzy and possibilistic transaction tables were extracted, retaining almost all knowledge hidden in these data with a clear reduction of their dimensionality. The proposed adaptation of the Apriori algorithm (initially designed for binary data) helps to extract frequent (fuzzy and possibilistic) itemsets, as well as the corresponding AR. This adaptation was based on a redefinition of the concepts of union and cardinality. This process, applied in primal and dual, discovered three types of biological relationships between genes, organs (areas of the body), and the developmental phases of the model species studied. Comparing the results of these two models shows that they discover AR of biological interest, especially the possibilistic approach, which is the most adequate to model the uncertainty in the studied data. MAP preserves better the boundaries between areas of overlapped gen-exp, and provides resistance to noisy pixels. Also, this modeling favors AR based on cases with clear gen-exp, to improve the quality and biological significance of extracted knowledge.

Modeling normal gen-exp during the phases of development of species will help to understand the causes of genetic diseases. To improve this work, future studies can explore modeling based on evidential logic (Dempster-Shafer theory) to redefine the union and cardinality of itemsets. In addition, sequential pattern mining can help extract sequential relationships between genes during different embryonic development phases.

REFERENCES

[1] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, "Spatiotemporal data mining: a survey on challenges and open problems," *Artificial Intelligence Review*, vol. 55, no. 2, pp. 1441–1488, Feb. 2022, https://doi.org/10.1007/s10462-021-09994-y.

[2] B. Karlik, "Soft computing methods in bioinformatics: A comprehensive review," *Mathematical and Computational Applications*, vol. 18, no. 3, pp. 176–197, 2013.

[3] L. Richardson *et al.*, "EMAGE mouse embryo spatial gene expression database: 2014 update," *Nucleic Acids Research*, vol. 42, no. D1, pp. D835–D844, Jan. 2014, https://doi.org/10.1093/nar/gkt1155.

[4] L. Ruzicka *et al.*, "The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources," *Nucleic Acids Research*, vol. 47, no. D1, pp. D867–D873, Jan. 2019, https://doi.org/10.1093/nar/gky1090.

[5] J. Thurmond *et al.*, "FlyBase 2.0: the next generation," *Nucleic Acids Research*, vol. 47, no. D1, pp. D759–D765, Jan. 2019, https://doi.org/10.1093/nar/gky1003.

[6] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Mar. 1993, pp. 207–216, https://doi.org/10.1145/170035.170072.

[7] T. Denœux, D. Dubois, and H. Prade, "Representations of Uncertainty in Artificial Intelligence: Probability and Possibility," in *A Guided Tour of Artificial Intelligence Research*, P. Marquis, O. Papini, and H. Prade, Eds. Cham: Springer International Publishing, 2020, pp. 69–117.

[8] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the Eleventh International Conference on Data Engineering*, Taipei, Taiwan, 1995, pp. 3–14, https://doi.org/10.1109/ICDE.1995.380415.

[9] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, Mar. 1996, pp. 1–12, https://doi.org/10.1145/233269.233311.

[10] K. Kianmehr, M. Kaya, A. M. ElSheikh, J. Jida, and R. Alhajj, "Fuzzy association rule mining framework and its application to effective fuzzy associative classification," *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 6, pp. 477–495, 2011, https://doi.org/10.1002/widm.40.

[11] Y. Djouadi, S. Redaoui, and K. Amroun, "Mining Association Rules under Imprecision and Vagueness: towards a Possibilistic Approach," in *2007 IEEE International Fuzzy Systems Conference*, London, Jun. 2007, pp. 1–6, https://doi.org/10.1109/FUZZY.2007.4295455.

[12] X. Geng, Y. Liang, and L. Jiao, "EARC: Evidential association rule-based classification," *Information Sciences*, vol. 547, pp. 202–222, Feb. 2021, https://doi.org/10.1016/j.ins.2020.07.067.

[13] D. Gangaramani and R. Londhe, "Survey on association rule analysis: Exploration using mining analysis," *International Journal of Hybrid Intelligent Systems*, vol. 21, no. 1, pp. 2–13, Feb. 2025, https://doi.org/10.3233/HIS-240015.

[14] J. van Hemert and R. Baldock, "Mining Spatial Gene Expression Data for Association Rules," in *Bioinformatics Research and Development*, 2007, pp. 66–76, https://doi.org/10.1007/978-3-540-71233-6_6.

[15] A. Houari, W. Ayadi, and S. B. Yahia, "Mining Negative Correlation Biclusters from Gene Expression Data using Generic Association Rules," *Procedia Computer Science*, vol. 112, pp. 278–287, Jan. 2017, https://doi.org/10.1016/j.procs.2017.08.262.

[16] R. Vengateshkumar, S. Alagukumar, and R. Lawrance, "Boolean Association Rule Mining on Microarray Gene Expression Data," in *Advanced Computing and Intelligent Engineering*, 2020, pp. 99–111, https://doi.org/10.1007/978-981-15-1081-6_9.

[17] S. H. Bouazza, "A Deep Ensemble Gene Selection and Attention-guided Classification Framework for Robust Cancer Diagnosis from Microarray Data," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20235–20241, Feb. 2025, https://doi.org/10.48084/etasr.9476.

[18] S. Acharya, S. Saha, and P. Pradhan, "Novel symmetry-based gene-gene dissimilarity measures utilizing Gene Ontology: Application in gene clustering," *Gene*, vol. 679, pp. 341–351, Dec. 2018, https://doi.org/10.1016/j.gene.2018.08.062.

[19] N. E. M. Khalifa, M. H. N. Taha, D. Ezzat Ali, A. Slowik, and A. E. Hassanien, "Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach," *IEEE Access*, vol. 8, pp. 22874–22883, 2020, https://doi.org/10.1109/ACCESS.2020.2970210.

[20] M. Ciortan and M. Defrance, "GNN-based embedding for clustering scRNA-seq data," *Bioinformatics*, vol. 38, no. 4, pp. 1037–1044, Jan. 2022, https://doi.org/10.1093/bioinformatics/btab787.

[21] T. Zeng, R. Li, R. Mukkamala, J. Ye, and S. Ji, "Deep convolutional neural networks for annotating gene expression patterns in the mouse brain," *BMC Bioinformatics*, vol. 16, no. 1, May 2015, Art. no. 147, https://doi.org/10.1186/s12859-015-0553-9.

[22] J. W. Bohland *et al.*, "Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy," *Methods*, vol. 50, no. 2, pp. 105–112, Feb. 2010, https://doi.org/10.1016/j.ymeth.2009.09.001.

[23] S. Papadimitriou, S. Mavroudi, and S. D. Likothanassis, "Mutual Information Clustering for Efficient Mining of Fuzzy Association Rules with Application to Gene Expression Data Analysis," *International Journal on Artificial Intelligence Tools*, vol. 15, no. 02, pp. 227–250, Apr. 2006, https://doi.org/10.1142/S0218213006002643.

[24] L. A. Zadeh, "Fuzzy sets, information and control," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[25] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3–28, Jan. 1978, https://doi.org/10.1016/0165-0114(78)90029-5.

[26] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 2020.

[27] D. Dubois and H. Prade, "Possibility Theory," in *The Palgrave Encyclopedia of the Possible*, Palgrave Macmillan, 2022, pp. 1064–1073.

[28] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Databases*, 1994, pp. 487–499.

[29] C. M. Kuok, A. Fu, and M. H. Wong, "Mining fuzzy association rules in databases," *SIGMOD Rec.*, vol. 27, no. 1, pp. 41–46, Nov. 1998, https://doi.org/10.1145/273244.273257.

[30] D. Dubois, E. Hüllermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowledge Discovery*, vol. 13, no. 2, pp. 167–192, Sep. 2006, https://doi.org/10.1007/s10618-005-0032-4.

[31] M. Pereyra, "Maximum-a-Posteriori Estimation with Bayesian Confidence Regions," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 285–302, Jan. 2017, https://doi.org/10.1137/16M1071249.

[32] D. Szklarczyk *et al.*, "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic Acids Research*, vol. 51, no. D1, pp. D638–D646, Jan. 2023, https://doi.org/10.1093/nar/gkac1000.