Expert Systems With Applications 237 (2024) 121593

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multi-instance learning with application to the profiling of multi-victim homicides

Rosario Delgado [a,*], Héctor Sánchez-Delgado [b]

[a] Department of Mathematics. Universitat Autònoma de Barcelona, Edifici C- Campus de la UAB. Av. de l'Eix Central s/n., 08193 Bellaterra (Cerdanyola del Vallès), Spain

[b] Data Quality and Statistics Executive. Kantar, Carrer de Can Calders, 4, 08173, Sant Cugat del Vallès, Spain

ARTICLE INFO

Keywords:
Bayesian Network
Combination rule
Criminal profile
Ensemble
Multi-instance (MI) learning
Multiple/mass homicide

ABSTRACT

Homicide involving multiple victims has a significant negative effect on society. Criminal profiling consists of determining the traits of an unknown offender based on those of the crime and the victims, with a view to their identification. To provide the most likely profile of the perpetrator of a multi-victim homicide, we propose a predictive model of supervised machine learning based on a Bayesian Network. Conventional classifiers can generate the perpetrator's profile according to the traits of each of the victims of the same homicide, but the profiles may differ from one another. To address this issue, we consider the *Multi-Instance* (MI) *learning* framework, in which the victims of the same incident form a bag, and each bag is associated with a unique label for each of the perpetrator's features. We introduce the ***unanimity MI assumption*** in this domain, and accordingly allocate a label to the bag based on the labels and probabilities the Bayesian Network has assigned its instances, using a combination rule from those of the *ensemble* of classifiers. We apply this methodology to the Federal Bureau of Investigation (FBI) homicide database to compare three combination rules empirically in the validation process, as well as theoretically, using the one that ultimately proves to be the best to build the final model, which is then applied in some illustrative examples to achieve the criminal profile.

1. Introduction

Due to its brutality and the fear it instills, homicide is one of the most serious sorts of crime, being also one of the main causes of death in the United States of America (U.S.), following Fowler, Leavitt, Betz, Yuan, and Dahlberg (2021). We must shed light and clarify the statistics to comprehend the scope of homicides in the U.S. According to data from the Federal Bureau of Investigation (FBI), there were 14,244 homicides on average per year from 2000 to 2021, with a rate of 4.58 per 100,000 inhabitants. To put it in a global context, following the United Nations Office on Drugs and Crime, in 2017 the homicide rate per 100,000 inhabitants was much lower in the vast majority of European countries, such as Spain, with 0.7, the United Kingdom with 1.2 or France, with 1.3. On the other hand, the countries of the American continent have much higher rates in general. For example, Costa Rica had 12.18 and Mexico, 25.71. So, we could say that on a global scale, the U.S. has a medium homicide incidence rate, although the harsh reality is that if we compare the U.S. with other developed countries, its rate is significantly higher.

Murder and manslaughter are two distinct legal offenses that are included in "homicide", as explained in Newburn (2013). The distinction between them is that murder implies that the victim was intended to be killed with malice aforethought, while manslaughter is the lesser form of homicide, which happens when death results even if there was no conscious desire to murder. Many U.S. states distinguish between two types of manslaughter: voluntary and involuntary, the first corresponding to the deliberate killing of another person during a sudden altercation, when they were provoked and were unable to control the situation, when they had some sort of diminished capacity due to a mental health issue, or when they had the sincere but ultimately irrational belief that using deadly force was necessary for self-defense. A person who kills another person unlawfully by criminal negligence, that is, without intending to do so, is said to have committed involuntary manslaughter.

In addition, we can divide homicides into four large groups, attending to the number of victims per event or incident (time-place combination), and the number of events per perpetrator: single, multiple/mass, spree and serial. Single and multiple/mass homicides occur in a single event, the first with a single victim, the multiple with 2 or 3 victims, and the mass with more than 3. The difference between serial murder and the multiple/mass homicide is that although in both cases there are multiple victims, the latter occurs in a single act while serial

* Corresponding author.
E-mail addresses: Rosario.Delgado@uab.cat (R. Delgado), Hector.Sanchez@kantar.com (H. Sánchez-Delgado).

https://doi.org/10.1016/j.eswa.2023.121593
Received 6 April 2023; Received in revised form 18 August 2023; Accepted 10 September 2023
Available online 15 September 2023

murder occurs at separate times and places, including a period of inactivity. This differentiation was established by FBI agent Robert Ressler in 1974, who also coined the term "serial killer". Following Sutton and Keatley (2021), the spree murder also kills different victims in different locations, but there is no cool-off period between them. Specifically, in this work we are interested in developing a new methodology that can be used to study multi-victim homicides (both, murders and voluntary manslaughter, with multiple/mass victims), committed in the same time and place and registered as a single incident or event. We apply this methodology to the case study of the FBI homicide database, which covers a period of 35 years, from 1980 to 2014, with data from all 51 states of the union.

Regardless of its incidence in terms of frequency or number of cases, multi-victim homicide is a crime with a high impact on society that can be considered a persistent public health problem in the U.S. (see Fowler et al., 2021). Hence, any instrument that can assist in identifying the profile of the perpetrator of such a homicide will be valuable for both law enforcement personnel responsible for upholding the law, and for legislators who design targeted policies aiming at eliminating this social blight.

Criminal profiling

In general, criminal profiling is a technique that makes it possible to identify the traits of a criminal in relation to their crimes and victims, the patterns of relationships between homicide victims and their killer being of the utmost interest. It is frequently used to try to provide details about a serial killer that could aid in their identification and capture (Kocsis, 2006). But it can also be used for other types of homicides, as we do in this work. We will not attempt to develop a profile for a specific series of murders carried out by the same person. Instead, we will develop a predictive model that will give investigators the most likely profile of the killer in the event of a multi-victim homicide. According to Fox (2022), homicides make up nearly a quarter of the crimes appeared in the offender profile publications, demonstrating the interest in profiling this kind of offender, who is also the criminal most typically portrayed in movies and television.

Criminal profiling is generally based solely (or mainly) on the expertise of investigators and forensic psychologists, and will then inevitably be subject to cultural biases or prejudices, inaccuracies, and misinterpretations (see Kocsis, 2006; Palermo & Kocsis, 2004; Turvey, 2002), although lately it is evolving towards a more rigorous process, adding geographic analysis techniques, for example. To avoid the weaknesses of this classic approach, machine learning methodologies can be used to build objective and quantitative decision making tools to support profiling investigations, as in Brahan, Lam, Chan, and Leung (1998), Gottschalk (2006) and Strano (2004). This is possible today by taking advantage of the opportunity to access the impressive computing power of current computers, and criminal databases of solved crimes, such as the one used in this work, which is the FBI database of solved homicides in the U.S.

Our objective is to develop a Machine Learning tool (a knowledge-based expert system) to predict certain characteristics of the perpetrator from those of the homicide victims, learned from a real database, which can help profile perpetrators of multi-victim homicides (considering that this type of homicide represents a challenge compared to single-victim homicides, as we will discuss later). Specifically, this tool will be a *probabilistic classifier* that learns the dependency relationships between these features and uses them to make the most plausible predictions for those that are unknown among them, which will be those corresponding to the perpetrator of the crime. As the profile of the aggressor consists of different characteristics (output variables) to be predicted, that is, different variables or criteria according to which to classify the perpetrator (age, race, …), we have to predict the value of more than one target variable, which may or may not be correlated (multi-label classification). To do this, we use a Bayesian Network (BN from now on), which is a probabilistic classifier that has the capacity to capture the interdependence relationships between all the variables of the model, including the output variables, and that allows considering more than one as output variables to be predicted from the rest.

Multi-instance (MI) classification

The usual classifiers can provide a profile of the perpetrator according to the characteristics of each of the victims of the multi-victim homicide, but the profiles may differ from each other. This poses a major problem as they have to give a single profile altogether. Somehow, we have to get them to "agree" on what their murderer looks like, and that is the challenge we face in this research. As a solution, we propose to consider the victims of the same incident as forming a bag. In this way, it could be considered that we move in the *multi-instance* (MI) *classification* environment described in Alpaydin, Cheplygina, Loog, and Tax (2015), in which each bag is associated with a unique label (for each of the output variables).

Multi-instance (MI) *learning* was introduced by Dietterich, Lathrop, and Lozano-Pérez (1997) for drug activity prediction, and is generally applied in contexts where data is formed in terms of bags or sets of instances, these bags being the ones that carry the labels and not so the individual instances. In this setting, the *standard* MI *assumption* (see Alpaydin et al., 2015; Carbonneau, Cheplygina, Granger, & Gagnon, 2018; Foulds & Frank, 2010), which assumes a binary classification task, states that the labels of the bags are determined by the disjunction of the labels of the instances in the manner described below: each instance has a hidden class tag that can be either + or −, and + is assigned to a bag if and only if one or more instances that make up the bag are + .

The objective of the paper is to build a predictive model to assign the labels at the bag level, with the particularity that it cannot be done directly, but first the predictive model must be built at the instance level and, later, the label will be assigned to each bag from the labels and probabilities assigned to the instances that compose it. For what reason? Because each victim contributes an instance or case to the database, providing information on each of the variables. That is, each victim determines a case him/herself and cannot be considered as an incomplete part of a whole, as in the *standard MI assumption*.

Related works

In general, the *standard MI assumption* is useful in situations where it is appropriate to describe an object by a set of parts, each of which is an instance carrying only a portion of the information needed for classification. This is the case, for example, of visual recognition, where the image to be recognized is usually divided into small fragments, or in text categorization, where the instances correspond to the small paragraphs into which the text is divided. But even in these cases, the MI *assumption* is not guaranteed to hold and is generally too restrictive to handle real-life situations, as explained by Küçükasci and Baydogan in Küçükasci and Baydogan (2018), where encoding strategies are introduced to represent the bags using the frequency of the instances in each of the pieces into which the feature space is partitioned. Some other authors have relaxed this assumption by allowing other different interactions between the labels of the instances and the label assigned to the bag they belong to. For example, Alpaydin et al. (2015) consider that some of the instances can carry a label but it is not known which of the bag can, which forces the MI approach to be adapted to a situation halfway between the pure instance-level approach and the bag-level classification. In addition, to address a problem of detecting failures in industrially manufactured entities, Graur, Maris, Potolea, Dinsoreanu, and Lemnaru (2018) introduce two approaches to solve MI *learning* problems in which the MI *assumption* is not met. We join these authors in their heterodox vision of MI, but in a different situation from the one they consider, so we will have to make different premises, in line with Foulds and Frank (2010). These premises are grouped under the name of *unanimity MI assumption* below.

On the other hand, there are different scientific papers in which a knowledge-based approach of the issue of criminal profiling is carried out. What we will do now is briefly comment on some of the works carried out by other authors, without claiming to be exhaustive, to highlight the differences with ours. The first works, to our knowledge, on the application of BN to criminal profiling are those of Baumgartner, Ferrari, and Salfati (2005) and Baumgartner, Ferrari, and Palermo (2008). In them, a BN is generated from a dataset of single-victim homicides, but previously the collection of alternative structures to consider is restricted using expert knowledge. The database is tiny (a few hundred cases) despite the big amount of variables (36 victim characteristics plus 21 criminal traits), and contrast to the K-fold cross-validation technique we use, the validation is carried out by split-validation, which prevents a statistical study to compare the suggested model with others. Despite this flaw, these works remain to be a reference for the use of the BN methodology for profiling. This methodology has also been applied by Delgado, González, Sotoca, and Tibau (2016, 2018) to the study of profiles of forest arsonists to understand their motivation, constructing archetypes that help identify the culprits, based on the data provided by the Spanish government.

Using homicide data recorded for one year in the National Incident-Based Reporting System U.S. database, Yang and Olafsson (2011) construct various machine learning classifiers, including Decision Trees, Random Forests, Support Vector Machines and Neural Networks, to predict the relationship between murder victims and offenders. They are limited to the context of single-label classification since they only consider one variable in the murderer's profile, which is the relationship with the victim. Instead, we also forecast other aspects of the perpetrator (multi-label classification). Furthermore, they carry out their single-label multi-class classification by dividing it into several simpler binary classification tasks. The point is that their solution cannot always be properly translated to the original multi-class one. However, BNs allow multi-label multi-class classification to be carried out naturally, without the need of any artifice. Finally, the authors also use split-validation, which does not allow for a robust statistical study to adequately compare the four classifiers.

Another interesting study is presented in Fowler et al. (2021), looking at multi-victim homicides in the U.S. With the help of bivariate statistical methods like chi-square tests and post-hoc pairwise comparisons with Bonferroni corrections, as well as Exploratory Data Analysis, the authors looked at the characteristics of mass, multiple, and single-victim homicides. Their findings are based on a database of homicides from 35 of the 51 states, which covers a period of 15 years (from 2003 to 2017). Therefore, this study's spatio-temporal window is visibly narrower than ours. Yet, they have variables that we do not, such as the incident's location, which is a very valuable information.

To wrap up this succinct overview, we note that Pecino-Latorre, Pérez-Fuentes, and Patró-Hernández (2019) utilize the Classification and Regression Trees (CART) methodology to determine the characteristics of homicides using a database of 448 homicide cases in Spain. The authors develop a different model for each of the six target output variables they consider, thus missing the chance to improve the model's predictive power by taking use of potential dependencies between them. In other words, they employ the simplest method for solving a multi-label classification –*binary relevance*– which ignores any potential relationship between labels. The Bayesian Networks methodology that we have used in our work enables to not overlook these dependencies.

Paper contributions

The main contributions of this paper are as follows:

1. In line with the specific problem under investigation, which is criminal profiling of homicides with multiple victims, we present two relevant conditions tailored to the context of MI *learning* that we have jointly called **unanimity MI assumption**. This assumption is applicable to any problem domain corresponding to events identified with bags, each of them composed of several individuals or elements. Each individual is identified with an instance or case, and all those involved in the same event are part of the corresponding bag. The conditions are:
   - (i) the instances cannot be considered as part of a whole (each of them does not cover an aspect or part of the event in which they are involved, but the total),
   - (ii) each instance is assigned a label, but with the restriction that the labels of all instances that make up a bag must match, hence the term "unanimity" in its name. The same label is then assigned to the bag.
2. We built a predictive Machine Learning model leveraging the methodology of the combination rules of the *ensembles* of classifiers to predict the profile of a perpetrator involved in a multiple homicide. With this approach, a single classifier is used to predict the characteristics of the perpetrator based on the attributes of the victims. To implement the classification following the *unanimity* MI *assumption* introduced in the previous item, we present a predictive Machine Learning model that has been designed following these two steps:
   - (a) *Classifier construction*: since each instance has a label assigned to it, this step can and is done at the instance level by learning a classifier as in single-instance learning, bypassing that the instances are grouped into bags.
   - (b) *Prediction*: Once the classifier is built, the prediction and its validation procedure will be carried out at the bag level, since the objective is the labeling of the bags. To do this, we use the labels and probability distributions assigned by the classifier to each of the instances that make up any new (unseen) bag, to assign it a label using a suitable algorithm. This algorithm is a combination rule (or scheme) specific to the *ensemble* of classifiers (explained in Section 2.3 below).

   Our approach is original, since focusing on multi-victim homicides, it applies an innovative Supervised Machine Learning methodology that combines a probabilistic classifier (specifically, a Bayesian Network, although others could be used as well) at the instance level, with the MI *learning* approach combined with *ensembling*. The details will be revealed throughout the work.
3. Application of the Predictive Model to the FBI homicide database: the predictive model developed is rigorously applied to the FBI's comprehensive database of criminal homicides. This real-world application demonstrates the model's ability to analyze and predict homicidal offender profiles based on victim-related features. By deploying the model with real data, the paper underscores its practicality and relevance within the domains of law enforcement and criminal profiling.

In summary, the paper makes specialized assumptions for MI *learning*, creates a new machine learning predictive model that employs a combination rule for the ensemble of classifiers, and validates the model's performance by applying it to the FBI homicide database. These contributions collectively advance the understanding and application of machine learning methodologies to predict perpetrator characteristics in multiple homicide situations.

The organization of the rest of the paper is as follows: in Section 2 we present the data set, the Bayesian Networks as probabilistic classifiers and the *ensembling* of combiners to be used in the sequel, as well as the validation procedure. Section 3 shows the results we have obtained, both for the Bayesian Network as predictive model, and for the prediction at the bag level, comparing three combination schemes through the different output variables, using the performance

Table 1
The 16 variables in the final database. (1) and (2) are used to exclude some cases and then removed (Fig. 1) while (3) is used to construct the new variable Bag, which is an auxiliary, and is then removed. Variables marked with (d) have been discretized, while those marked with (r) have been recoded. Variables in blue are those of the victim, variables of crime are in teal, and those of the perpetrator are in red, including their relationship to the victims.

Variables in the original database			Incorporated
Removed	Mantained	Modified	
Record.ID	State	Victim.Age$^{(d)}$	Bag
Agency.Code	Year	Victim.Race$^{(r)}$	
Agency.Name	Month	Victim.Count$^{(r)}$	
Agency.Type	Victim.Sex	Perpetrator.Age$^{(d)}$	
City	Victim.Ethnicity	Perpetrator.Race$^{(r)}$	
Crime.Solved$^{(1)}$	Perpetrator.Sex	Perpetrator.Count$^{(r)}$	
Crime.Type$^{(2)}$	Perpetrator.Ethnicity	Weapon$^{(r)}$	
Record.Source		Relationship$^{(r)}$	
Incident$^{(3)}$			

metrics that we have introduced. We discuss our findings and conclude in Section 4, while the appendices include some tables (Appendix A) and figures (Appendix B). In Appendix C can be found the pseudo-code for the algorithms that implement the combination schemes and a comparison of their accuracy in the binary case.

2. Materials and methods

2.1. The data set

Data source

We got the FBI data through the kaggle repository.[1] The most comprehensive U.S. homicide database currently available, this database includes murders from the FBI's Supplementary Homicide Report and covers a 35-year period, from 1980 to 2014 inclusive. It is compiled by the Murder Accountability Project, founded by Thomas Hargrove, to whom we wish to thank. Initially, the data included information on 24 variables for 638,454 cases (now and hereafter, a "case" or "instance" is identified with a victim).

Exclusion criteria

Different criteria have been used to exclude cases and variables, until forming the final database with 36,269 cases and 16 variables. Regarding the cases, Fig. 1 explains how the selection of those that make up the final database has been carried out. Regarding the variables, some have been eliminated, others have been modified and a new one has also been created and incorporated. We list them in Table 1 and give details in the *Data description* subsection below.

Data description

Now we describe the variables in the database (Table 1) and, in particular, we explain the preprocessing phase in which we have modified eight of them, six by recoding, and two by discretization. We also explain the construction of the auxiliary variable Bag, which has been incorporated into the final database. Table A.11 shows the relative frequencies (in percentages) of each category of the different

[1] https://www.kaggle.com/datasets/murderaccountability/homicide-reports?resource=download.

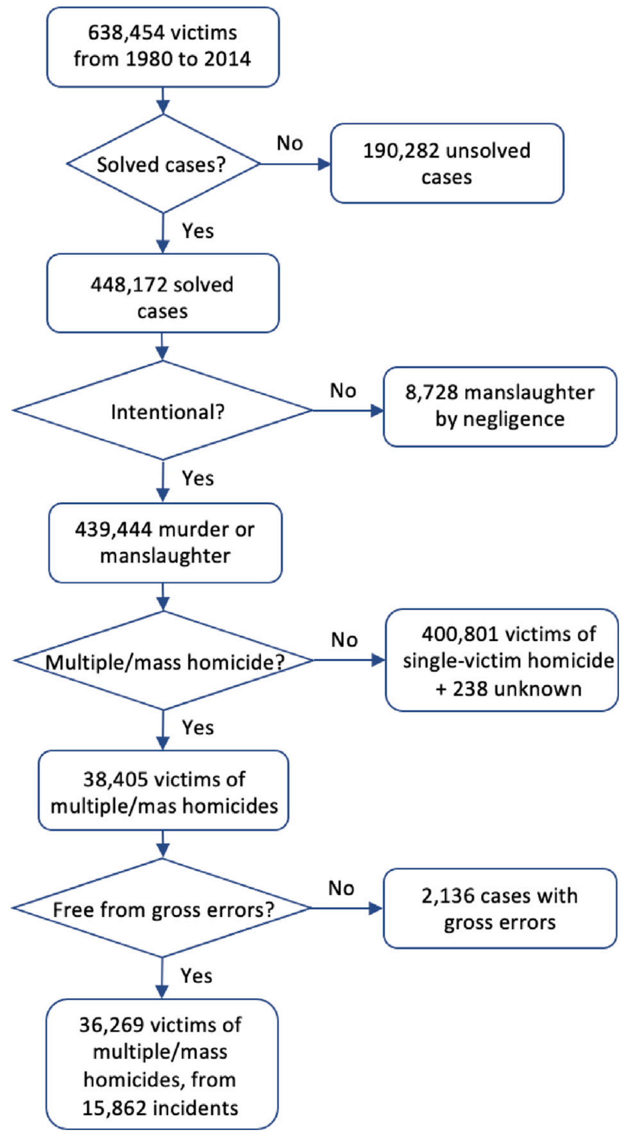

Fig. 1. Flowchart of the exclusion of victims from the initial database. To exclude unsolved cases we use the "Crime.Solved" variable in the original database, while to exclude victims of manslaughter by negligence we use the "Crime.Type" variable.

variables, distinguishing between the homicides with multiple (2 or 3) and massive (4 or more) victims.

State: The 51 states of the union contribute cases to the database, with California contributing the most (5300) and South Dakota the least (40), in absolute terms.

Year: The data covers a 35-year period, from 1980 to 2014, peaking in 1992 with 1307 victims and reaching the absolute minimum in 2014, with just 844.

Month: The distribution of cases per months is quite balanced, varying between 2722 in November and 3305 in January.

Sex: The vast majority of the perpetrators are men (94.0% male, 6.0% female), with which the database is very unbalanced with respect to the variable `Perpetrator.Sex`, while as regards the victims, the variable `Victim.Sex` is much more balanced, with 61.4% male, and 38.6% female. This fact evidences the sociological reality that women continue to be victims much more frequently than perpetrators of serious crimes, especially homicides with various victims.

Ethhicity and Race: The terms *race* and *ethnicity* are mutually independent, according to the U.S. Office of Management and Budget definition adhered to by the U.S. Census Bureau. Regardless of race, *ethnicity* makes a distinction between people who claim to have ancestral origins in Spain or Latin America (Hispanic or Latino), and those who do not (non-Hispanic).[2] In 2020, 62 million of residents in the U.S., or 18.7% of the population, identified as Hispanic, of whom 12.5 million or 20.3% self-identified as White only[3] down from the 2019 American Community Survey, when 38.3 million, or 65.5% of Latinos self-identified as White.[4] In the datasets at hand, the vast majority of individuals categorized as Hispanic are also White.

Specifically, of the 3784 victims registered as Hispanic in our final database, 3723 are know to be White (98.4%), while of the 10,255 non-Hispanics, 6348 are registered as White (61.9%). From the 1643 Hispanic perpetrators, 1616 (98.4%) are recorded as White, and from the 4558 perpetrators known to be non-Hispanic, 2565 are recorded as White, representing 56.3%. However, for the reasons mentioned above, we have kept *ethnicity* and *race* as two distinct variables, for both victims and perpetrator, rather than merging them into one. However, our predictive model will capture the actual dependency between them.

We have grouped race into three groups: White, Black (or African American), and a third category, Native-Islander, which includes both Asian and Pacific Islander on the one hand, and Native American and Alaska Native on the other (this grouping has been motivated by the scarcity of individuals of each of these races in the database, which together represent 3.55% of the perpetrators, and 3.79% of the victims).

Age: Regarding age, it has been discretized into ranges or intervals, according to a criterion of approximate equity between them, taking into account that the frequency in the central intervals will naturally be higher than in the extremes, and distinguishing the minority of legal age (18 years), but with some differences that we want to highlight between the authors and the victims. In both cases, ages 99 years and over have been coded as "Unknown". The main difference is that we have also recoded as "Unknown" the ages of the perpetrators in case they are less than 10 in the original database. Although this is arbitrary, it has been motivated by the fact that some of the ages were suspiciously low, being 0 years in many cases, which suggests an error in data collection. Furthermore, due to the scarcity of cases in those ages, for the aggressor we have merged into a single category the intervals from 50 to 64 years, and 65 or more, which remain as different intervals for the age of the victims. It is logical that the perpetrator's age moves towards the central age ranges with respect to that of the victims, having low frequencies in the lower range (10–17, 6.7%) and in the upper range ($\geq 50, 8.6\%$).

Count: Despite their names, `Perpetrator.Count` and `Victim.Count` are variables with very different meanings in the FBI database. Indeed, whereas `Victim.Count` is, for any victim, the total number of victims of the same incident discounting this victim, and has been recoded as 1, 2 (multiple homicide) and >2 (mass homicide), for any perpetrator, `Perpetrator.Count` appears to be, although it is not perfectly clear from the database documentation provided by the FBI, the number of previous incidents in which the perpetrator had been involved, and has been recoded as binary with values 0 and 1, the latter being the category that groups any number of previous incidents greater than 0.

Weapon: Our intention was to group some of the categories of this variable, corresponding to the means used by the perpetrator to commit the homicide, by similarity, if they were rare. But due to its great variability, we have only done in these two cases: we have grouped in `Long_hung` both `rifle` and `shotgun`, and we have merged `Drugs` and `Poisson` in a single category. Finally, we have 12 categories (plus the `Unknown`).

Relationship: Since it influences the type of aggressiveness, the relationship between the perpetrator and the victims of an homicide has been considered as a cornerstone in the majority of studies on homicides, as in Yang and Olafsson (2011). In this work we have grouped this variable into four categories: `IPV` (Intimate Partner Violence), `family` (homicide committed by relatives, including in-laws, which cannot be considered IPV, such as, for example, son, daughter, father, mother, sister, brother, …), `aquaintance` (where the perpetrator is a friend, neighbor, employee or employer, …) and `stranger` (committed by a person who does not know or is known by the victims). Our classification is slightly different from that of Yang and Olafsson (2011), where the category `Close to family` includes other homicides in addition to `IPV`, which we have considered as a category by itself, due to its special social relevance and criminological interest, although it only applies to the 12.3% of the perpetrators of our database. The rest of categories correspond to `stranger` (22.4%), `family` (25.9%) and `acquaintance` (39.4%), the percentages corresponding to the perpetrators for whom the relationship to the victims was known.

Bag: This auxiliary variable has been introduced as an essential tool to uniquely identify the cases (victims) of the same event, all of them with the same perpetrator, assigning them a distinct number. There are 15,862 different bags (referred to different incidents or perpetrators), with 4236 victims mass homicide victims killed in 858 incidents (average of 4.9 victims/incident), and 32,033 multiple homicide victims killed in 15,004 incidents (average of 2.1 victims/incident).

2.2. The probabilistic classifier

As mentioned in the Introduction, we will use a probabilistic classifier to predict each of the output variables for each of the victims that are part of the same incident (bag), and apply a combination rule to obtain the joint prediction for the bag. Of the different methodologies used in Machine Learning to learn classifiers from a database of solved cases, we consider *probabilistic classifiers* because they not only predict the label, but also estimate the probability distribution over the set of possible labels. As such a classifier we build a Bayesian Network with all the model variables except `Bag`, both input variables corresponding to the event: `Year`, `Month`, `State` and `Weapon`, and to the characteristics of the victim: `Sex`, `Age`, `Race`, `Ethnicity` and `Count`, as well as the output, which are the characteristics of the murderer: `Sex`, `Age`, `Race`, `Ethnicity`, `Count` and `Relationship` (see Table 1).

The probabilistic relationships between the variables that affect a phenomenon – in this case, a multi-victim homicide – can be represented graphically by BNs (Koller & Friedman, 2009), which are graphical models utilized for probabilistic inference. A standard BN is a model that depicts the joint probability distribution P of a set of random variables (discrete or categorical), which are represented graphically as nodes in a *directed acyclic graph* (DAG). The **Markov condition** asserts that each node in the DAG is independent of those nodes that are not its descendants, provided that the values taken by its parents are known. Directed arcs between nodes entail conditional (not necessarily causal) dependencies that are governed by this condition.

The updating of probabilities with the BN from a specific piece of evidence is referred to as *Bayesian Inference*; we compute *a posteriori*

[2] Nancy L. Fisher (1996). Cultural and Ethnic Diversity: A Guide for Genetics Professionals. Johns Hopkins University Press. p. 19. ISBN 978-0-8018-5346-3. https://books.google.es/books?id=mqXlA7e4VN8C&pg=PA19&redir_esc=y#v=onepage&q&f=false.

[3] https://www.census.gov/data/tables/2020/dec/2020-redistricting-supplementary-tables.html.

[4] https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/.

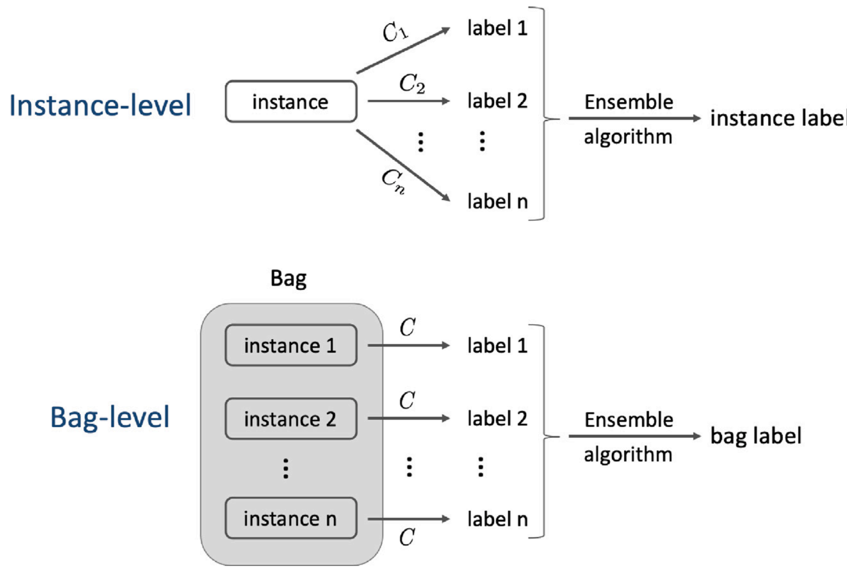

Fig. 2. Scheme of classification differences at the instance and bag levels. $C_1, \ldots, C_n$ are the n different classifiers ensembled at the instance level, while C is the only one classifier used at the bag level, whose predictions for the n instances are ensembled.

probabilities from the evidences and the *a priori* probabilities. According to the Maximum A Posteriori (MAP) criterion, the prediction of a query variable given an evidence is its instantiation with the highest *a posteriori* probability, and this probability is referred to as the *confidence level* (CL) of the prediction. We employ the Maximum Likelihood Estimation (MLE) approach to estimate the parameters, which are the conditional probabilities of any variable to its parents, and the marginal probabilities of the root nodes, which are the nodes without parents in the DAG. With the restriction that no directed arcs are allowed from any input variable to any output variables, we learn the structure of the BN using the Hill-Climbing Score-Based Structure Learning algorithm, which provides a pseudo-optimal DAG that (locally) maximizes the Bayesian Information Criterion (BIC) score. For this we use the `hc` function from the R package `bnlearn` by Scutari (2010). Learning and prediction algorithms have been implemented in the R language (R Core Team, 2013), with the help of `bnlearn`.

2.3. Ensembling for the MI learning with the unanimity MI assumption

Ensemble of classifiers (also known as "combined classifiers") is a method that has been widely applied in classification learning to create a new classifier by merging a set of base classifiers in accordance with a rule or scheme.

The algorithm that we propose here follows the same line, but instead of combining, given an instance, the predictions that different classifiers provide for it, what it does is use a single classifier and combine the predictions that it provides for the different instances that make up each of the bags, with which the predictions at the bag level are obtained, as schematically represented in Fig. 2. In fact, we will use an ensemble algorithm to combine the predictions given by the classifier for the different instances that form a bag (the victims of the same incident/perpetrator), for any target variable. To do this, we must choose the combination scheme that we will apply. We consider three different rules (more details can be found in Appendix C, see also Delgado (2022)):

1. *Majority Vote* MV: the most well-known, straightforward, and widely used **hard voting** rule, which selects the label with the highest number of votes from the predicted for each instance of the bag.
2. *Ensemble Average* EA: from the perspective of predictive power, it is the most effective **soft voting** rule, choosing the label with the highest average of the predicted probabilities that the classifier assigns for each instance of the bag.
3. *Confidence Level based Majority Vote* CL-MV: a **semi-hard voting** rule halfway between MV and EA. Each label has a counter that sums the predicted probability assigned to that class by the classifier for all the instances in the bag for which that label is the predicted one (and 0 otherwise). CL-MV is just going with the prediction for the label that maximizes the counter. If n is odd and the label predicted for each instance is right with a high enough probability, this combination strategy performs better in the binary classification setting ($r = 2$) when compared to MV, according to Delgado (2022), where in a sensitivity analysis, it has also been compared with EA and found to be significantly more resilient to probability estimation error.

Note that in the context of the *unanimity* MI *assumption*, it makes no sense to use trainable combination schemes that have additional parameters needing to be trained, such as those that give different weights to different predictions. This is due, in particular, to the fact that we do not know a priori what the number of instances forming a bag is, nor do we have a clear criteria to prioritize some instances over others, which would serve to assign them different weights.

2.4. Validation procedure

We choose to perform a K*-fold cross-validation* procedure with $K = 10$ (see Bishop) to validate and compare the different predictive models. In this way, we obtain a confusion matrix for any model and any $k = 1, \ldots, K$, and from them we can calculate the following measures of performance: *accuracy* and Matthews Correlation Coefficient (MCC), introduced in the binary case by Matthews (1975), and extended to the multi-class setting by Gorodkin (2004), for all the output variables.

The Mean Absolute Error (MAE) (see Cardoso and Sousa (2011)) is applied for ordinal classification when `Perpetrator.Age` is predicted, which is a categorical variable of a quantitative nature that has been discretized into 5 intervals: $[10, 18)$, $[18, 25)$, $[25, 35)$, $[35, 50)$, ≥ 50.

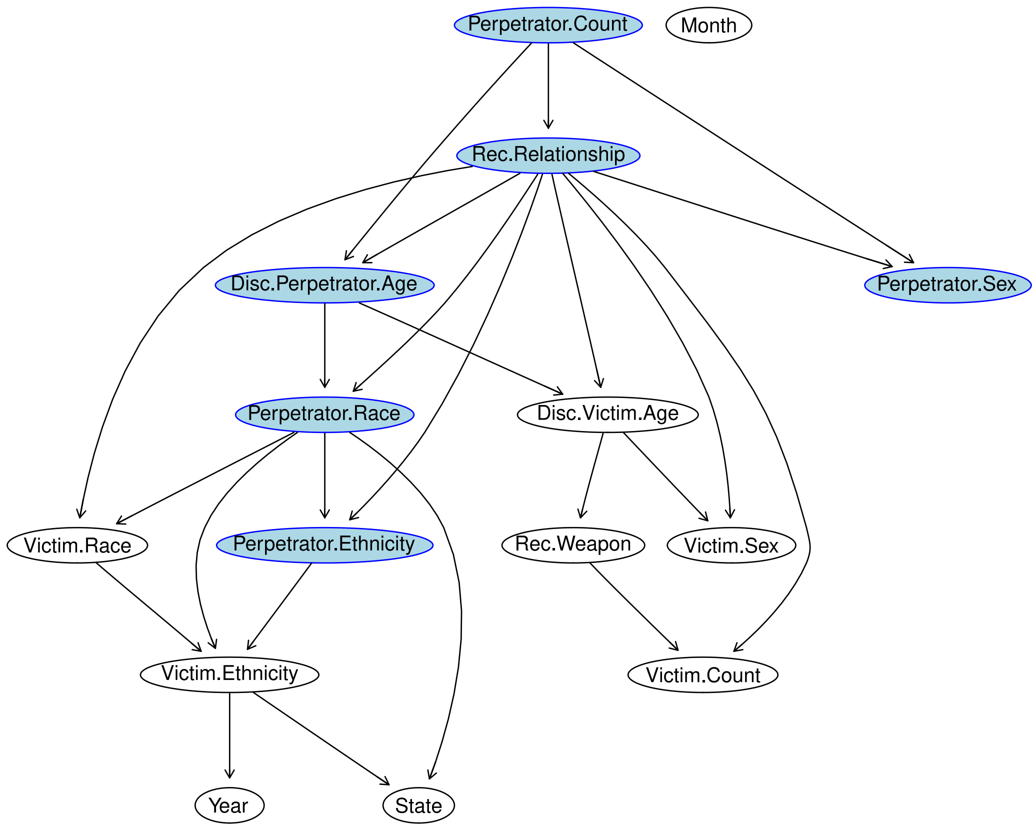

Fig. 3. DAG of the Bayesian Network used as classifier C for the prediction at the instance level, before the use of the combination schemes for the *unanimity MI assumption*. In blue the nodes corresponding to the output (perpetrator) variables.

3. Results

3.1. The Bayesian network

In Fig. 3 we show the relationships between the characteristics of the victim and the crime, on the one hand, and the perpetrator, on the other, captured by the DAG of the BN learned from the entire database, at the instance level.

We can use the `arc.strength` function from the `bnlearn` package to obtain a measure of the strength of the probabilistic relationships expressed by the directed arcs of the BN, and obtain that all arcs have a very significant value of strength. We can establish which variables play the main role in the model using measures of centrality and/or betweenness taken from the area of Network Analysis applied to the DAG in Fig. 3.

Graph's important (influential) nodes are identified using centrality indicators in the fields of Graph Theory and Network Analysis, where "importance" here is defined as a measure of node's contribution to the cohesiveness of the network. We adopt the *degree of centrality of Freeman* (Freeman, 1977) and the *basic standard betweenness measure* by Borgatti and Everett (2006) as indicators. Their values, which are calculated for each variable, are provided in Table 2.

The variables highlighted in Table 2: `Relationship`, `Victim.Ethnicity` and `Perpetrator.Race`, followed by `Victim.Age` and `Perpetrator.Age`, act as gateways, and the arcs connecting them as bridges through which information flows from one set of variables in the model to another.

3.2. Predicting at the bag level

In Table 3 we record the average of the $K = 10$ values obtained by K-fold cross-validation for any of the metrics, the output variables and the different combination rules for the ensemble used for the *unanimity MI assumption* (CL-MV, MV and EA). The corresponding boxplots are found in Figs. B.4–B.5. From a descriptive point of view, we can observe a clear advantage for the CL-MV ensemble combination scheme with all the metrics: *accuracy*, MAE and MCC.

The question now is whether these differences observed at a descriptive level are statistically significant or not. To figure it out we need to perform some statistical hypothesis testing. In this sense, Table 4 reports whether any of the combination schemes, the one registered in the row, is statistically significantly better (p-value < 0.10) than the one registered in the column, according to the different metrics used in the experiment. We use the Holm-Bonferroni correction for multiple comparisons to avoid detecting significant differences that are

Table 2
(Normalized to sum up 100) Freeman's degree of centrality and measure of basic standard betweenness for the variables in the model. In bold, the three most influential variables and their values of centrality and betweenness.

Variable	Freeman's centrality (%)	Betweenness (%)
State	4.16667	0
Year	2.08333	0
Month	0	0
Victim.Sex	4.16667	0
Victim.Race	6.250000	3.24074
Victim.Ethnicity	**10.416667**	**22.22222**
Victim.Age	8.33333	13.88889
Victim.Count	4.16667	0
Weapon	4.16667	5.55556
Perpetrator.Sex	4.16667	0
Perpetrator.Race	**12.500000**	**24.07407**
Perpetrator.Ethnicity	6.250000	3.24074
Perpetrator.Age	8.33333	6.94444
Perpetrator.Count	6.350000	0
Relationship	**18.750000**	**20.83333**

Table 3
Mean values of the metrics obtained by K-fold cross-validation for any of the output variables, using the different combination rules for the ensemble used in the *unanimity MI assumption* (CL-MV, MV and EA). By row, we indicate the order from **best (1)** to worst (3).

	Mean *accuracy*		
	CL-MV	MV	EA
Race	**0.6161285 (1)**	0.5476253 (3)	0.5551590 (2)
Ethnicity	0.5931035 (2)	0.5391634 (3)	**0.6120940 (1)**
Age	**0.3781464 (1)**	0.3596979 (2)	0.2307187 (3)
Count	**0.5851616 (1)**	0.5622917 (3)	0.5790462 (2)
Relationship	0.4398548 (2)	0.4189362 (3)	**0.4434323 (1)**
	Mean MAE		
	CL-MV	MV	EA
Age	**448.2 (1)**	460.5 (2)	634.1 (3)
	Mean MCC		
	CL-MV	MV	EA
Race	**0.3089283 (1)**	0.21792450 (3)	0.27643789 (2)
Ethnicity	0.2167372 (2)	0.0451777 (3)	**0.25091453 (1)**
Age	**0.1343040 (1)**	0.10409507 (2)	0.01568847 (3)
Count	0.1797634 (2)	0.13375927 (3)	**0.18797883(1)**
Relationship	**0.2286003 (1)**	0.19496783 (3)	0.22395685 (2)

Table 4
p-values for the comparison of the mean values of the metrics used, with the Holm-Bonferroni correction for multiple comparisons. Alternative hypothesis: the combination scheme of the row is **better than** that of the column.

Accuracy Row > Column	CL-MV	MV	EA
CL-MV		Race: 0.0017** Age: 0.016* Relationship: 0.066****	Race: 0.0063** Age: 2.5×10^{-7}***
MV			Age: 1.7×10^{-6}***
EA		Relationship: 0.079****	
MAE Row < Column	CL-MV	MV	EA
CL-MV		Age: 0.024*	Age: 5.2×10^{-9}***
MV			Age: 1.4×10^{-7}***
MCC Row > Column	CL-MV	MV	EA
CL-MV		Race: 0.012* Age: 0.00574** Relationship: 0.031*	Age: 9.9×10^{-6}***
MV			Age: 0.00023***

* Only significant p-values are recorded: at 5%.
** Only significant p-values are recorded: at 1%.
*** Only significant p-values are recorded: 1‰.
**** Denotes a slight significance, at 10%.

not truly significant, using the pairwise Wilcoxon signed-rank test or the Student's t-test to compare pairs of samples corresponding to the same run, as appropriate according to the Shapiro–Wilk normality test, which has been performed previously.

All in all, the CL-MV combination scheme is clearly the best of those considered, since none of the others surpassed it with any of the target variables. More specifically, from Table 4 we can state that there are no significant results for the variables `Count` or `Ethnicity`, and that for the rest of output variables:

- *Accuracy*: CL-MV outperforms the other combination schemes for the variables `Race` and `Age`, and outperforms MV for the variable `Relationship`.
- MAE (for the variable `Age`): CL-MV outperforms both MV and EA.
- MCC: CL-MV combination scheme shows to be better than MV for `Race` and `Relationship`, and better than both MV and EA for `Age`.

To evaluate the impact of various incident or victim factors on the forecasted profile of the unknown offender, we have thus decided, in light of these results, to employ the final predictive model built using the CL-MV combination scheme from the complete database.

An example of application

Just to give an example, imagining one particular incident out of the myriad of them that could occur in reality, let us consider that the incident is a **mass** homicide with 4 non-Hispanic victims, all under 10 years of age, a male (victim V2) and three females, who were killed with a firearm (`State`, `Year` and `Month` have not been included as input information about the incident). We carry out two sensitivity studies in this example, as an illustration of the potentiality of the model to generate knowledge.

Sensitivity study 1. Table 5 shows the predicted characteristics for the perpetrator obtained with our predictive model, using CL-MV, with their associated confidence level, if the race of the victims were changed, one by one, from 4 White to 4 Black.

We can observe how the predicted race for the perpetrator moves accordingly from White to Black, how the predicted age decreases and how the predicted relationship goes from "family" to "acquaintance" (the sex of the perpetrator is always predicted to be male, and so has been omitted, and the same goes for the variable `Count`, which is always predicted to be 0, and the variable `Ethnicity`, predicted to be non-Hispanic). Table 6 shows the confidence of the individual-level predictions for any of the incident victims, based on their race.

Algorithm 2 (Appendix C) with the CL-MV combination rule uses the information in Table 6 to derive the predicted perpetrator profile for the entire incident that is reported in Table 5. For example, in Setting 1.1, with respect to `Relationship`, three of the victims (`V1`, `V3` and `V4`) predict `Family` with the same confidence level 0.4127566, while the fourth predicts `Stranger` with 0.383445. Using Algorithm 2 with CL-MV to obtain the predicted label for the incident with the four victims, we compare the sum of the confidence levels assigned to `Family` by the victims: $3 \times 0.4127566 = 1.2382698$, with the corresponding sum assigned to `Stranger`: 0.383445, resulting in the prediction being `Family`, where $1.2382698/(1.2382698 + 0.383445) = 0.7635558$ is its confidence level, as recorded in Table 5.

Sensitivity study 2. Another possibility would be to study the evolution of the predicted profile of the perpetrator in the example, when the race of the four victims is White but their ethnicity changes, one by one, from non-Hispanic to Hispanic, as shown in Table 7. Now the predicted race of the perpetrator is always White, and then omitted. Table 8 shows, analogously to Table 6, the confidence of the predictions at the individual level for any of the victims of the incident, based on their ethnicity.

In the sensitivity study 2, to predict `Relationship` in Setting 2.3 (Table 7), for example, we observe in Table 8 that `V1` and `V2` predict `Stranger` with confidence levels 0.2959600 and 0.4218292, respectively, while the rest of the victims predict `Family`, with confidence 0.4127566 each. To use CL-MV in Algorithm 2 (Appendix C), we compare the sum of the confidence levels assigned to `Stranger`, which is 0.7177892, and the sum for `Family`, 0.8255132. The resulting prediction is the one corresponding to the largest, which is `Family`, with confidence level $0.8255132/(0.8255132 + 0.7177892) = 0.5349005$.

We observe from Tables 5, 7 that the predicted race/ethnicity of the perpetrator matches that of the majority of the victims, with the confidence level increasing as the number of victims of that race/ethnicity does the same. The predicted age of the perpetrator tends to be lower if the victims are Black than White, but if they are White, it is not affected by ethnicity. The predicted relationship changes from "family" to "acquaintance" if the victims' majority race is Black, and to "stranger" if their ethnicity is Hispanic (of race White). We can see that depending on the proportion of races and ethnicities in the group, this **mass**

Table 5
Sensitivity study 1: Example of predicted perpetrator profile using CL-MV, for an incident with four non-Hispanic victims (mass homicide), <10 years old, when the weapon is a firearm, as the race of the victims varies, one by one, from White to Black.

Mass homicide. **Sensitivity study 1**				
Victim		Predicted perpetrator profile (incident)		
Race		Race	Age	Relationship
Setting 1.1	V1. White, V2. White V3. White, V4. White	White 1	[25,35) 1	Family 0.7635558
Setting 1.2	V1. Black, V2. White V3. White, V4. White	White 0.7527732	[25,35) 0.7243044	Family 0.5589648
Setting 1.3	V1. Black, V2. Black V3. White, V4. White	White 0.5123915	[18,25) 0.5520587	Family 0.5737075
Setting 1.4	V1. Black, V2. Black V3. Black, V4. White	Black 0.7416863	[18,25) 0.7836611	Acquaintance 0.4140512
Setting 1.5	V1. Black, V2. Black V3. Black, V4. Black	Black 1	[18,25) 1	Acquaintance 0.699362

Table 6
Sensitivity study 1: Individual predictions for any of the four victims in the example (mass homicide), based on the victim's race. In parentheses, the corresponding confidence level of the individual prediction.

Mass homicide. **Sensitivity study 1**				
Victim	Race	Individual predicted perpetrator profile for each victim		
		Race	Age	Relationship
V1	White Black	White (0.8932712) Black (0.8646841)	[25,35) (0.2883954) [18,25) (0.3338198)	Family (0.4127566) Acquaintance (0.2679026)
V2	White Black	White (0.8463076) Black (0.8354481)	[25,35) (0.3002172) [18,25) (0.3770374)	Stranger (0.383445) Stranger (0.3454937)
V3	White Black	White (0.8932712) Black (0.8646841)	[25,35) (0.2883954) [18,25) (0.3338198)	Family (0.4127566) Acquaintance (0.2679026)
V4	White Black	White (0.8932712) Black (0.8646841)	[25,35) (0.2883954) [18,25) (0.3338198)	Family (0.4127566) Acquaintance (0.2679026)

Table 7
Sensitivity study 2: Analogous to the example in Table 5 with four White victims (mass homicide) as their ethnicities vary, one by one, from non-Hispanic to Hispanic.

Mass homicide. **Sensitivity study 2**				
Victim		Predicted perpetrator profile (incident)		
Ethnicity		Ethnicity	Age	Relationship
Setting 2.1	V1. non-Hispanic, V2. non-Hispanic V3. non-Hispanic, V4. non-Hispanic	Non-Hispanic 1	[25,35) 1	Family 0.7635558
Setting 2.2	V1. Hispanic, V2. non-Hispanic V3. non-Hispanic, V4. non-Hispanic	Non-Hispanic 0.7807124	[25,35) 1	Family 0.5485436
Setting 2.3	V1. Hispanic, V2. Hispanic V3. non-Hispanic, V4. non-Hispanic	Non-Hispanic 0.543309	[25,35) 1	Family 0.5349005
Setting 2.4	V1. Hispanic, V2. Hispanic V3. Hispanic, V4. non-Hispanic	Hispanic 0.7158239	[25,35) 1	Stranger 0.710652
Setting 2.5	V1. Hispanic, V2. Hispanic V3. Hispanic, V4. Hispanic	Hispanic 1	[25,35) 1	Stranger 1

Table 8
Sensitivity study 2: Individual predictions for any of the four victims in the example (mass homicide), based on the ethnicity of the victim. In parentheses, the corresponding confidence level of the individual prediction.

Mass homicide. **Sensitivity study 2**				
Victim	Etnicity	Individual predicted perpetrator profile for each victim		
		Etnicity	Age	Relationship
V1	Non-Hispanic Hispanic	Non-Hispanic (0.9290091) Hispanic (0.7783224)	[25,35) (0.2883954) [25,35) (0.2965019)	Family (0.4127566) Stranger (0.2959600)
V2	Non-Hispanic Hispanic	Non-Hispanic (0.9129823) Hispanic (0.7834779)	[25,35) (0.3002172) [25,35) (0.3035398)	Stranger (0.383445) Stranger (0.4218292)
V3	Non-Hispanic Hispanic	Non-Hispanic (0.9290091) Hispanic (0.7783224)	[25,35) (0.2883954) [25,35) (0.2965019)	Family (0.4127566) Stranger (0.2959600)
V4	Non-Hispanic Hispanic	Non-Hispanic (0.9290091) Hispanic (0.7783224)	[25,35) (0.2883954) [25,35) (0.2965019)	Family (0.4127566) Stranger (0.2959600)

Table 9
Example of predicted perpetrator profile using CL-MV, for an incident with three non-Hispanic victims (multiple homicide), <10 years old, when the weapon is a firearm, as the race of the victims varies, one by one, from White to Black.

Multiple homicide. **Sensitivity study 1**			
Victim	Predicted perpetrator profile		
Race	Race	Age	Relationship
V1. White, V2. White V3. White	White 1	[35,50) 0.6741045	Family 0.7402884
V1. Black, V2. White V3. White	White 0.6670902	[25,35) 0.6685955	Family 0.7046505
V1. Black, V2. Black V3. White	Black 0.655939	[18,25) 0.3615387	Family 0.6756024
V1. Black, V2. Black V3. Black	Black 1	[25,35) 0.6440647	Family 0.6264575

Table 10
Analogous to the example in Table 9 with three White victims (multiple homicide) as their ethnicity varies, one by one, from non-Hispanic to Hispanic.

Multiple homicide. **Sensitivity study 2**			
Victim	Predicted perpetrator profile		
Ethnicity	Ethnicity	Age	Relationship
V1. non-Hispanic, V2. non-Hispanic V3. non-Hispanic	Non-Hispanic 1	[35,50) 0.6741045	Family 0.7402884
V1. Hispanic, V2. non-Hispanic V3. non-Hispanic	Non-Hispanic 0.7057269	[25,35) 0.6579581	Family 0.6998363
V1. Hispanic, V2. Hispanic V3. non-Hispanic,	Hispanic 0.6245493	[25,35) 0.6593685	Family 0.6809751
V1. Hispanic, V2. Hispanic V3. Hispanic	Hispanic 1	[25,35) 1	Family 0.6239864

homicide could fit into the two scenarios described in the Discussion section in Fowler et al. (2021): homicides within the family, with a large number of minor victims (*scenario 1*) and homicides in public places such as schools, with most of the victims being young people and children, strangers or acquaintance to the perpetrator (*scenario 2*).

To quantify the association between the ethnic origin of the victims and the perpetrator, for example, we can calculate from Table 7 the Odds Ratio (OR) in favor of the non-Hispanic prediction for the perpetrator's ethnic origin, when the number of non-Hispanic victims is reduced from 3 (Setting 2.2) to 2 (Setting 2.3):

$$\mathrm{OR} = \frac{0.543309/(1-0.543309)}{0.7807124/(1-0.7807124)} = 0.3341546 \approx 0.33$$

that being less than 1 means that the reduction in the number of non-Hispanic victims reduces the odds in favor of the non-Hispanic ethnicity of the perpetrator. But it goes further, since OR quantifies this reduction in odds, and tells us that it happens to be a third, approximately. Just one more example: from Table 5, the OR in favor of a relationship of "acquaintance" when the number of Black victims goes from 3 (Setting 1.4) to 4 (Setting 1.5) is:

$$\mathrm{OR} = \frac{0.699362/(1-0.699362)}{0.4140512/(1-0.4140512)} = 3.29293 \approx 3.3$$

which means that the increase in the number of Black victims raises the odds in favor of "acquaintance", and quantifies this increase (more than triple).

What would have happened if instead of 4 victims, the incident had only 3 and was therefore a **multiple** homicide? Analogous to what we did in the mass homicide example, we now assume 3 non-Hispanic victims in the same incident, all under the age of 10, one male (victim V2) and two females, who were killed with a firearm. Tables 9–10 show the predicted perpetrator characteristics and associated confidence levels for the entire incident in two sensitivity studies, as the race or the ethnicity of the victims changes.

Comparing Tables 9–10 with Tables 5–7 we observe a higher predicted age in the case of the **multiple** homicides, and that the predicted relationship between victims and perpetrator is always "family", regardless of the number of White/Black and non-Hispanic/Hispanic victims, unlike **mass** homicides, for which the prediction was "acquaintance" when the number of Black victims is majority, or "stranger", when the majority of the victims were Hispanic. One of the two scenarios mentioned above, *scenario 1*, could be used to frame this multiple homicide.

4. Discussion and conclusions

In this study, we describe a knowledge-based expert system designed to assist in profiling homicide offenders, that is, to forecast their attributes (sex, age, race, ethnicity, count and relationship to victims), when there are multiple victims in a single incident. This profile is predicted based on the victim's characteristics (age, sex, race, ethnicity) and the incident's details (state, year, month and weapon), and it can be crucial early in the investigation so that law enforcement investigators can more accurately identify the unknown perpetrator. Rather than replace criminologists' capacity to profile suspects, smart software is intended to assist them in their goal of improving their effectiveness at preventing crime.

To address the issue of predicting the profile of the perpetrator of a multi-victim homicide, we adopt the *MI learning* paradigm. By definition, neither the pure instance-level approach, where instances are not considered as part of a bag for the classification task, nor the construction of a classifier exclusively at the bag-level without taking instances into account, can be used to solve this problem. Rather, the classifier, which in our case is a BN, must be built at the instance-level, and the labels and probabilities that have been assigned to each instance of the same bag must then be somehow combined to produce the tag for the bag as a whole. And this is where the combination schemes of the ensemble of classifiers methodology come into play, handling a MI problem that pure instance-level classifiers cannot address on their own.

The hybrid strategy known as CL-MV (semi-hard voting) has the best overall performance, according to our methodological comparison of the three combination rules: the soft-voting *majority vote* MV, the hard-voting *ensemble average* EA, and the hybrid approach CL-MV. Indeed, CL-MV performs better than the other for all the variables for which statistically significant differences are detected, with the performance metrics *accuracy* and MCC, in general, and MAE for the ordinal discretized variable `Age`. We also delved into the interpretability of the model, based on the DAG of the Bayesian Network from which it has been built. The most "influential" variables, as determined by several measures of centrality and betweenness, were found to be: `Relationship`, `Victim.Ethnicity` and `Perpetrator.Race`. The expert system not only enables profiling but also generates knowledge that is provided objectively and quantitatively. For instance, we can infer from the example presented in Tables 5–7 that the predicted relationship is more likely to be "acquaintance" if the victims are primarily Black, or "stranger" if they are predominantly Hispanic, as opposed to "family", if the victims are White and non-Hispanic. The Odds Ratio, which is obtained from the probabilities estimated with the model, can be used to quickly determine the strength of the association between the input and the output variables.

We believe that in some settings, such as multi-victim homicides, the *unanimity MI assumption* is an effective approach. We have revealed that the CL-MV criterion, as a rule to combine the labels and probabilities assigned to the instances in the same bag, has produced encouraging results when applied to the FBI homicide database in the U.S. The predictive model may self-adapt to other police databases, from other countries, or to other sorts of crimes with multiple victims, such as assaults, robberies, kidnappings, etc. because it was built with a Machine Learning methodology that learns from the database from which it feeds. Although the topic of predicting the profile of the perpetrator in a multi-victim homicide is a particular case, we are certain that this work will have more useful applications. For instance, predicting the qualities of a teacher or school based on the traits of the students, who make up a bag. Perhaps those of the creator of a piece of art when determining its authorship, or those of a sports coach based on the characteristics of the team members, ...

Ultimately, this research may be valuable because we developed an expert system that could be implemented to make it easier for law enforcement officers to use it for practical applications, such as early prevention of criminal recidivism, since aiding in the elaboration of criminal profiling will facilitate to identify and apprehend the culprits, preventing them from committing another homicide. We concur with the sentiment expressed in the Discussion of Fowler et al. (2021) in that preventing a mass homicide, especially if it involved a shooting, could have positive effects on preventing further crime, due to its potential contagion effect for almost two weeks (see also Towers, Gomez-Lievano, Khan, Mubayi, and Castillo-Chavez (2015)). At the same time, it also appears to be a solid starting point for upcoming criminological inquiries. Comparing the forecasted traits of the offender before and after the implementation of a legal measure that might have an effect on homicides might be instructive, for example.

The local data-driven machine learning methodology for determining a perpetrator's profile in a multi-victim homicide described in this paper, seems to be a useful and promising tool with significant criminological applicability insofar, as it can assist criminologists to create incident-tailored profiles, and authorities to better manage available resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to express their gratitude to the reviewers who, with their comments, helped them to make an improved final version of their work. We confirm that the manuscript has been written, read, and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

Funding

R. Delgado is supported by Ministerio de Ciencia e Innovación, Gobierno de España, Spain, project ref. PID2021-123733NB-I00

Appendix A. Data set: table of relative frequencies (in %)

See Table A.11

Appendix B. Additional figures

See Figs. B.4 and B.5

Table A.11
Relative frequencies (%) of the characteristics of the victim, of the perpetrator, and of the `weapon`, depending on whether the homicide is multiple or massive. Note that for the perpetrator, the age ranges 50–64 and ≥65 have been merged into ≥50, and that the minimum age is 10. It is remarkable, though not unexpected, that while comprising merely about 6% of the offenders, women constitute over 37% of the victims.

Characteristics of		Victim		Perpetrator	
		Multiple	Mass	Multiple	Mass
Sex	Female	37.54%	46.29%	6.07%	5.28%
	Male	62.46%	53.71%	93.93%	94.72%
Age	<10	9.21%	21.28%		
	10–17	8.58%	12.16%	6.66%	7.10%
	18–24	21.33%	15.81%	29.52%	28.29%
	25–34	23.23%	17.84%	31.32%	30.21%
	35–49	19.94%	18.28%	24.02%	24.68%
	50–64	10.81%	9.31%	8.48%	9.72%
	≥65	6.90%	5.32%		
Race	Native-Islander	3.64%	5.00%	3.56%	3.45%
	Black	32.49%	28.12%	38.19%	37.70%
	White	63.87%	66.88%	58.25%	58.85%
Ethnicity	Hispanic	27.51%	21.82%	26.75%	24.45%
	Non-Hispanic	72.49%	78.18%	73.25%	75.55%
Count (victim)	1	81.20%	0.00%		
	2	18.80%	0.00%		
	>2	0.00%	100.00%		
Count (perpetrator)	0			76.12%	76.53%
	1			23.88%	23.47%
Relationship	Acquaintance			39.38%	39.88%
	Family			25.87%	25.87%
	IPV			12.25%	12.90%
	Stranger			22.50%	21.35%

Weapon	Multiple	Mass	Weapon	Multiple	Mass
Blunt Object	4.47%	4.58%	Firearm	6.63%	5.95%
Drowning	0.32%	0.40%	Handgun	54.95%	54.98%
Drugs-Poisson	0.32%	0.00%	Knife	11.28%	10.41%
Explosives	0.29%	0.34%	Long hung	15.60%	17.39%
Fall	0.01%	0.06%	Strangulation	0.97%	0.57%
Fire	4.12%	4.52%	Suffocation	1.01%	0.80%

Appendix C. The combination schemes

For the sake of completeness, the pseudo-code of the algorithms for the implementation of the combination schemes MV, EA and CL-MV are below (see Delgado (2022)). To predict an output variable at the bag level for a bag composed of n instances $E_1, \dots, E_n$, first, we need to obtain the prediction for each of the instances by Algorithm 1, whose output feeds the others.

Algorithm 1 `Instance prediction`

Input instances $E_1, \dots, E_n$ forming a bag, classifier C, output variable classes $y_1, \dots, y_r$

Output the predicted classes y_j^* for any instance E_j, and the probability distributions $\{p_{jk},\ k = 1, \dots, r\}$, $j = 1, \dots, n$

1: **for** j in $1 : n$ **do**
2: **for** k in $1 : r$ **do**
3: **compute** p_{jk} (probability assigned by C to y_k, given evidence E_j)
4: $j_{max} = \arg\max_{k=1,\dots,r} p_{jk}$
5: $y_j^* = y_{j_{max}}$ (predicted class for instance E_j with the MAP criterion)
return $\{y_j^*,\ j = 1, \dots, n\}$ and $\{p_{jk},\ k = 1, \dots, r,\ j = 1, \dots, n\}$

Note that Algorithm 1 enables the prediction of any output variable for each instance using the Maximum A Posteriori (MAP) criterion. This involves assigning the class with the maximum probability. Subsequently, Algorithm 2 presented below facilitates the combination of predictions acquired for individual instances within a bag, using a

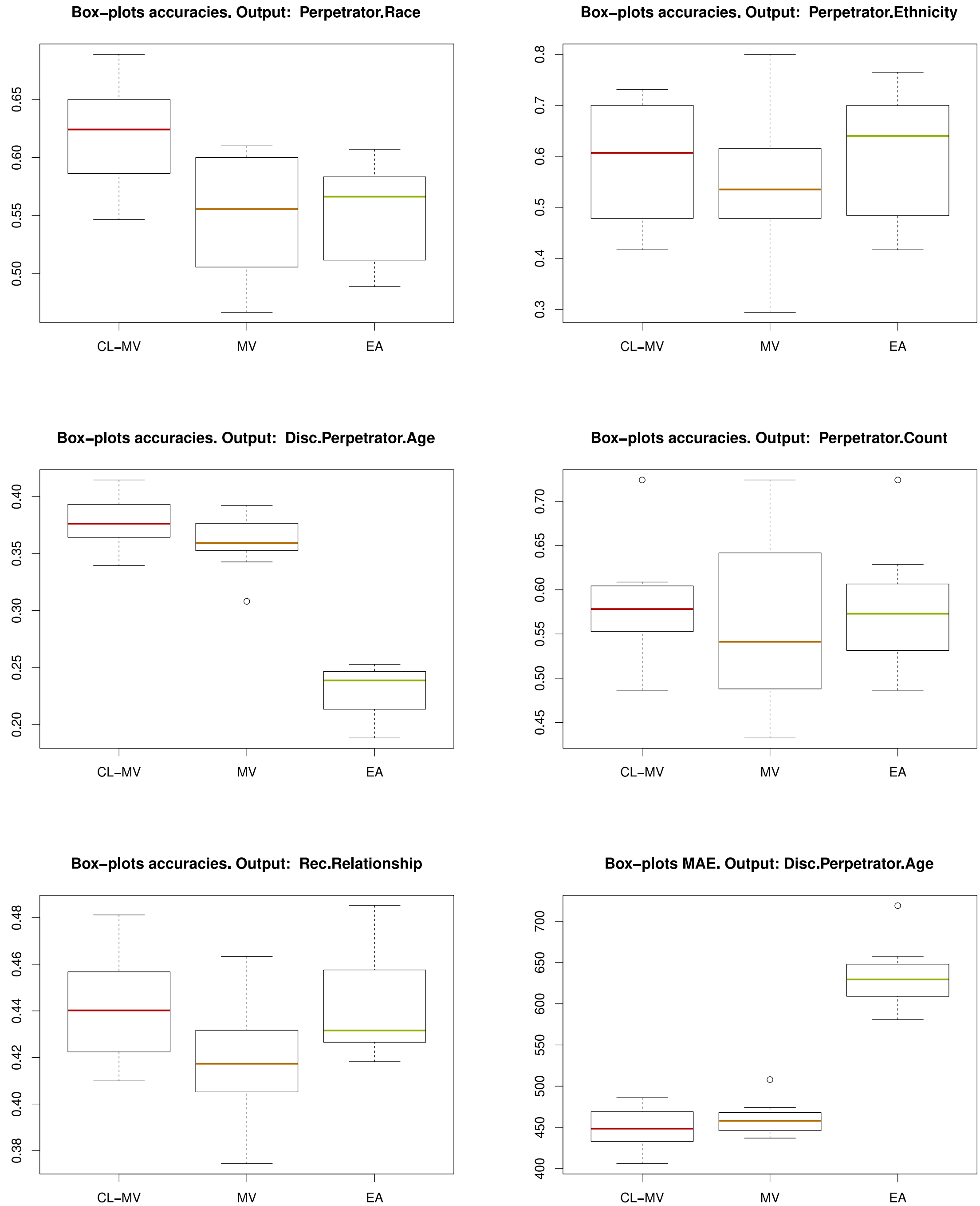

Fig. B.4. Boxplots for the accuracy (the higher the better) of the ensembles at the bag level, using the different combination schemes: CL-MV, MV, and EA, for each of the output variables, from the data obtained in the validation procedure, and boxplot of MAE for the ordinal output variable `Perpetrator.Age` (the smaller the better). Consistent with Table 4, CL-MV shows greater accuracy compared to MV across Race, Age and Relationship variables, and compared to EA across Race and Age. Notably, neither MV nor EA outperforms CL-MV across any output variables. Concerning Age, CL-MV exhibits significantly lower MAE than EA and slightly lower than MV.

specific combination rule. where `combination` $\in$ {MV, EA, CL-MV} and function g_k in step 3 of Algorithm 2 is $\frac{1}{n}\sum_{j=1}^{n} p_{jk}$ if `combination` = EA, and otherwise,

$$g_k = \begin{cases} 0 & \text{if } y(k) = \emptyset \\ \#y(k) & \text{if } y(k) \neq \emptyset,\ \text{for}\ \texttt{combination} = \text{MV} \\ \sum_{j\in y(k)} p_{jk} & \text{if } y(k) \neq \emptyset,\ \text{for}\ \texttt{combination} = \text{CL-MV}. \end{cases}$$

In words, EA selects the predicted class for the bag by maximizing the average (or sum) of the probabilities assigned to each class by the classifier across all n instances comprising the bag. On the other hand, MV designates the bag's class based on the most frequently chosen class among the instances. Meanwhile, CL-MV assigns the bag's prediction by maximizing the sum of probabilities assigned to each class by the classifier, but only considering the instances, out of n, that opted for that particular class.

Remark. MV offers the benefit that once we know the prediction of any of the instances, we do not need to maintain any other information, which is advantageous from a computational perspective and saves storage space. EA is more computationally and storage-demanding because it must save and use every probability distribution

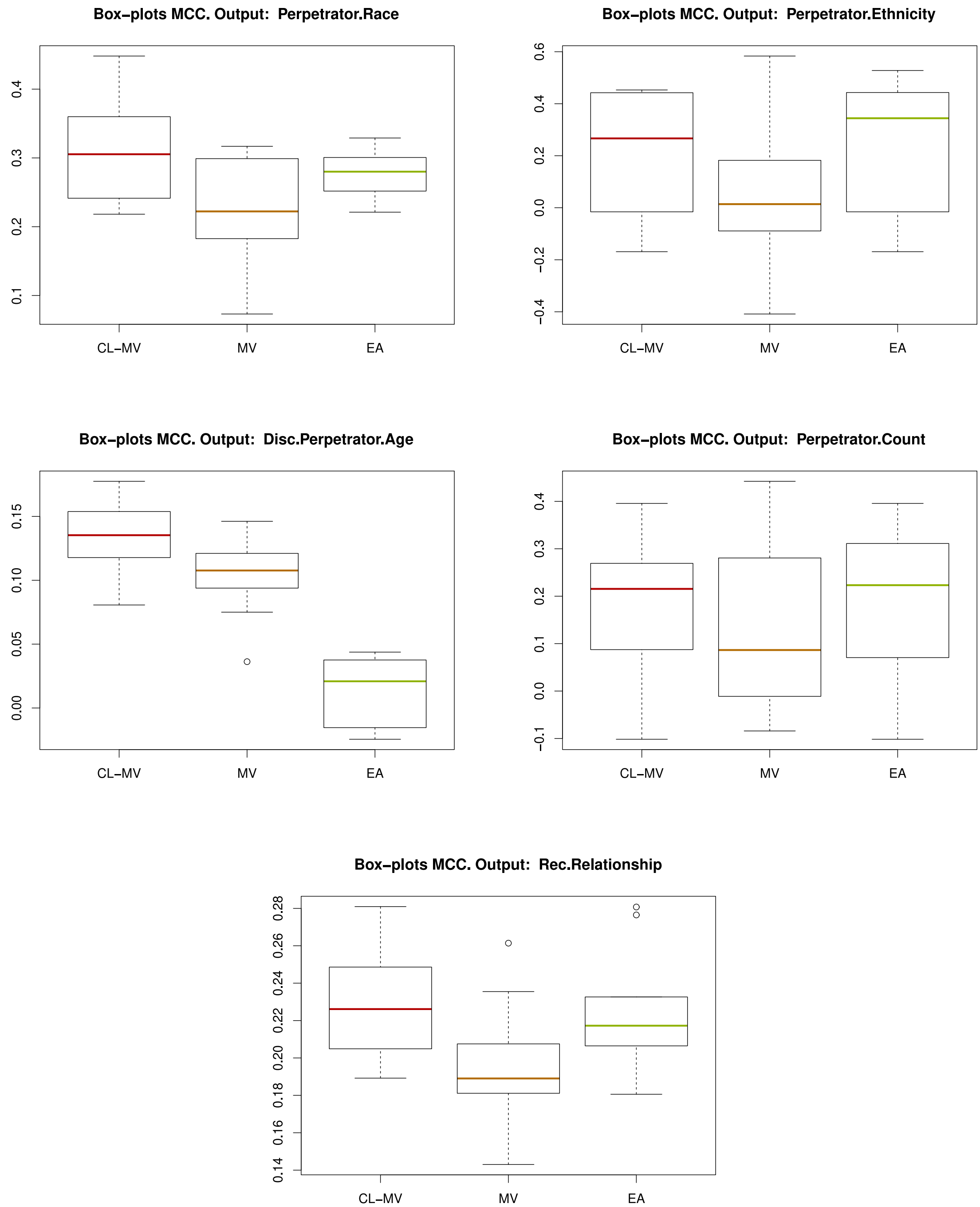

Fig. B.5. Boxplots for the MCC of the ensembles at the bag level, with combination schemes: CL-MV, MV and EA, for the output variables. The higher the better. As indicated by Table 4, CL-MV demonstrates greater MCC than MV for Race, Age and Relationship, and also outperforms EA for Age. Notably, neither MV nor EA outperforms CL-MV in terms of MCC across the output variables.

value assigned by the classifier to the class prediction for any instance. Halfway between them since only uses the maximum of the probability distribution for any instance, is CL-MV.

Comparing the combination schemes through *accuracy* in the binary case

Following Delgado (2022) we compare in the binary case the *accuracy* of the different combination rules we are considering, with an odd number of instances forming the bag, if the predictions given by the classifier for any of the instances forming a bag are right or wrong independently of each other. Let $p \in (0, 1)$ denote the probability that the classifier provides the correct class label for any of the instances. *Accuracy* for the different combination schemes is (see Delgado, 2022):

$$Acc_{\mathrm{MV}} = \sum_{\ell=\lfloor n/2 \rfloor+1}^{n} \binom{n}{\ell} p^{\ell} (1-p)^{n-\ell},$$

$$Acc_{\mathrm{CL\text{-}MV}} = \sum_{\ell=\lfloor (1-p)\, n \rfloor+1}^{n} \binom{n}{\ell} p^{\ell} (1-p)^{n-\ell},$$

$$Acc_{\mathrm{EA}} = p \quad \text{(independently of } n\text{, by construction)},$$

where $\lfloor x \rfloor$ denotes the integer part or *floor* of x. In Delgado (2022) the following result is stated and proved, and we reproduce it here adapted to our *unanimity MI assumption* context.

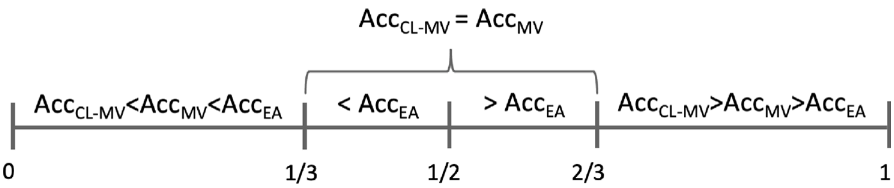

Fig. C.6. Scheme of the statement of Proposition 2: comparing the *accuracy* of the different combination schemes in the binary case ($r = 2$) with $n = 3$, for $p \in (0, 1)$.

Algorithm 2 `Combination scheme` (Algorithms 1–3 Delgado, 2022)

Input the output of Algorithm 1
Output the predicted class $y^*_{\texttt{combination}}$ for the bag

for k in $1 : r$ **do**
  $y(k) = \{j = 1, \dots, n \ : \ y^*_j = y_k\}$ (indices of instances with predicted class y_k)
  g_k
$\ell = \arg\max_{k=1,\dots,r} g_k$ (the class that maximizes function g)
$y^*_{\texttt{combination}} = y_\ell$
return $y^*_{\texttt{combination}}$

Proposition 1 (*Proposition 2 Delgado, 2022*). *In the binary case and with an odd number of instances giving rise to a bag, n, we have that*

$$\begin{cases} \text{If} \quad p \le \frac{n-1}{2n}, & \text{then} \quad Acc_{CL\text{-}MV} < Acc_{MV}, \\ \text{If} \quad \frac{n-1}{2n} < p \le \frac{n+1}{2n}, & \text{then} \quad Acc_{CL\text{-}MV} = Acc_{MV}, \\ \text{If} \quad p > \frac{n+1}{2n}, & \text{then} \quad Acc_{CL\text{-}MV} > Acc_{MV}. \end{cases}$$

It can be seen (Condorcet Jury Theorem, 1785, as reproduced in Shapley & Grofman, 1984) that

(a) if $p > 0.5$, $\lim_{n\to\infty} Acc_{MV} = 1$ and it is monotonically increasing,
(b) if $p = 0.5$, $Acc_{MV} = 0.5 = Acc_{EA}$ for all n,
(c) if $p < 0.5$, $\lim_{n\to\infty} Acc_{MV} = 0$ and it is monotonically decreasing,

that is, while for $p = 0.5$ MV and EA have the same *accuracy*, we can expect MV to show improvement over EA only if $p > 0.5$.

In the particular case $n = 3$, to be more specific in a concrete example, we have that

$$Acc_{\text{MV}} = 3\,p^2\,(1-p) + p^3\,,$$

$$Acc_{\text{CL-MV}} = \begin{cases} p^3 & \text{if } p \le \frac{1}{3} \\ 3\,p^2\,(1-p) + p^3 & \text{if } \frac{1}{3} < p \le \frac{2}{3} \\ 3\,p\,(1-p)^2 + 3\,p^2\,(1-p) + p^3 & \text{if } p \ge \frac{2}{3}\,. \end{cases}$$

The following result establishes the comparison between them, and shows that CL-MV is better (it has greater *accuracy*) than MV, and both are better than EA, when the probability that each instance of the bag gives the correct label, p, is large enough. Its proof is a simple exercise and is therefore omitted, while a schematic of the result can be seen in Fig. C.6.

Proposition 2. *In the binary case and with $n = 3$, we have that*

$$\textit{CL-MV vs. MV} \quad \begin{cases} \text{If} \quad p \le \frac{1}{3}, & \text{then} \quad Acc_{CL\text{-}MV} < Acc_{MV}, \\ \text{If} \quad \frac{1}{3} < p \le \frac{2}{3}, & \text{then} \quad Acc_{CL\text{-}MV} = Acc_{MV}, \\ \text{If} \quad p > \frac{2}{3}, & \text{then} \quad Acc_{CL\text{-}MV} > Acc_{MV}. \end{cases}$$

$$\textit{MV vs. EA} \quad \begin{cases} \text{If} \quad p < \frac{1}{2}, & \text{then} \quad Acc_{MV} < Acc_{EA}, \\ \text{If} \quad p = \frac{1}{2}, & \text{then} \quad Acc_{MV} = Acc_{EA}, \\ \text{If} \quad p > \frac{1}{2}, & \text{then} \quad Acc_{MV} > Acc_{EA}. \end{cases}$$

References

Alpaydin, E., Cheplygina, V., Loog, M., & Tax, D. M. J. (2015). Single- vs. multiple-instance classification. *Pattern Recognition*, *48*, 2831–2838.

Baumgartner, K. C., Ferrari, S., & Palermo, G. (2008). Constructing Bayesian Networks for criminal profiling from limited data. *Knowledge-Based Systems*, *21*, 563–572.

Baumgartner, K. C., Ferrari, S., & Salfati, C. G. (2005). Bayesian network modeling of offender behavior for criminal profiling. In *Proceedings of the 44th IEEE conference on decision and control* (pp. 2702–2709). Seville, Spain: http://dx.doi.org/10.1109/CDC.2005.1582571.

Bishop, C. M. *Series: Information science and statistics*, *Pattern recognition and machine learning*. NY: Springer New York, ISBN: 978-0-387-31073-2, Hardcover. Published: 17 August 2006.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*, 466–484.

Brahan, J. W., Lam, K. P., Chan, H., & Leung, W. (1998). AICAMS: Artificial intelligence, crime analysis and management system. *Knowledge-Based Systems*, 355–361.

Carbonneau, M. A., Cheplygina, V., Granger, E., & Gagnon, G. (2018). Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, *77*, 329–353.

Cardoso, J., & Sousa, R. (2011). Measuring the performance of ordinal classification. *International Journal of Pattern Recognition and Artifical Intelligence*, *25*(8), 1173–1195.

Delgado, R. (2022). A semi-hard voting combination scheme to ensemble multi-class probabilistic classifiers. *Applied Intelligence*, *52*, 3653–3677, https://doi-org.are.uab.cat/10.1007/s10489-021-02447-7.

Delgado, R., González, J. L., Sotoca, A., & Tibau, X. A. (2016). A Bayesian Network profiler for wildfire arsonists. In P. Pardalos, P. Conca, G. Giuffrida, & G. Nicosia (Eds.), *Lecture notes in computer science*: *vol. 10122*, *Machine learning, optimization and big data. MOD 2016* (pp. 379–390). Cham: Springer, http://dx.doi.org/10.1007/978-3-319-51469-7_31.

Delgado, R., González, J. L., Sotoca, A., & Tibau, X. A. (2018). Archetype of wildfire arsonists: An approach by using Bayesian Networks. In Janusz Szmyt (Ed.), *Forest fire*. IntechOpen, http://dx.doi.org/10.5772/intechopen.72615.

Dietterich, T., Lathrop, R., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, *89*, 31–71.

Foulds, J., & Frank, E. (2010). A review of multi-instance learning assumptions. *The Knowledge Engineering Review*, *25*(1), 1–25.

Fowler, K. A., Leavitt, R. A., Betz, C. J., Yuan, K., & Dahlberg, L. L. (2021). Examining differences between mass, multiple and single-victim homicides to inform prevention: findings from the National Violent Death Reporting System. *Injury Epidemiology*, *8*, 49, 15 pages.

Fox, B. (2022). Chapter 18 - Offender profiling: a review of the research and state of the field. In Paulo Barbosa Marques, & Mauro Paulino (Eds.), *Police psychology* (pp. 381–394). Academic Press, ISBN: 9780128165447, http://dx.doi.org/10.1016/B978-0-12-816544-7.00018-8.

Freeman, L. C. (1977). A set of measures of centrality based upon betweenness. *Sociometry*, *40*, 35–41.

Gorodkin, J. (2004). Comparing two k-category assignments by a k-category correlation coefficient. *Computational Biology and Chemistry*, *28*(5–6), 367–374.

Gottschalk, P. (2006). Stages of knowledge management systems in police investigations. *Knowledge-Based Systems*, *19*, 381–387.

Graur, D. O., Maris, R. A., Potolea, R., Dinsoreanu, M., & Lemnaru, C. (2018). Complex localization in the multiple instance learning context. In A. Appice, & et al. (Eds.), *LNAI*: *vol. 10785*, *NFMCP 2017* (pp. 93–106). Springer International Publishing AG.

Kocsis, R. N. (2006). *Criminal profiling: Principles and practice*. Totowa, NJ: Humana Press.

Koller, D., & Friedman, N. (2009). *Adaptive computation and machine learning*, *Probabilistic graphical models: Principles and techniques*. MIT Press, ISBN: 9780262013192.

Küçükasci, E. S., & Baydogan, M. G. (2018). Bag encoding strategies in multiple instance learning problems. *Information Sciences*, *467*, 559–578.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442–451.

Newburn, T. (2013). *Criminology* (2nd ed.). Oxon/New York: Routledge.

Palermo, G., & Kocsis, R. N. (2004). In Charles C. Thomas (Ed.), *Offender profiling: An introduction to the sociopsychological analysis of violent crime*. Springfield, IL.

Pecino-Latorre, M. M., Pérez-Fuentes, M. C., & Patró-Hernández, R. M. (2019). Homicide profiles based on crime scene and victim characteristics. *International Journal of Environmental Research and Public Health*, *16*, 3629. http://dx.doi.org/10.3390/ijerph16193629, 13 pages.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org/.

Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, *35*(3), 1–22, http://www.jstatsoft.org/v35/i03/.

Shapley, L., & Grofman, B. (1984). Optimizing group judgement accuracy in the presence of interdependencies. *Public Choice*, *43*, 329–343.

Strano, M. (2004). A neural network applied to criminal psychological profiling: an Italian initiative. *International Journal of Offender Therapy and Comparative Criminology*, *48*(4), 495–503.

Sutton, M. R., & Keatley, D. (2021). Cooling-off periods and serial homicide: A case study approach to analysing behaviour between murders. *Forensic Science International: Mind and Law*, *2*, Article 100066.

Towers, S., Gomez-Lievano, A., Khan, M., Mubayi, A., & Castillo-Chavez, C. (2015). Contagion in mass killings and school shootings. *PLoS ONE*, *10*(7), Article e0117259.

Turvey, B. E. (2002). *Criminal profiling: An introduction to behavioral evidence analysis* (2nd ed.). Academic Press.

Yang, R., & Olafsson, S. (2011). Classification for predicting offender affiliation with murder victims. *Expert Systems with Applications*, *38*, 13518–13526.