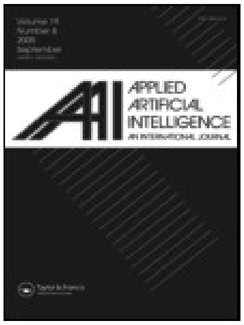


Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

DESIGNING EFFICIENT ANN CLASSIFIERS FOR MATCHING BURGLARIES FROM DWELLING HOUSES

Mohammad Reza Keyvanpour , Mohammad Reza Ebrahimi & Mostafa Javideh

To cite this article: Mohammad Reza Keyvanpour , Mohammad Reza Ebrahimi & Mostafa Javideh (2012) DESIGNING EFFICIENT ANN CLASSIFIERS FOR MATCHING BURGLARIES FROM DWELLING HOUSES, Applied Artificial Intelligence, 26:8, 787-807, DOI: 10.1080/08839514.2012.718227

To link to this article: https://doi.org/10.1080/08839514.2012.718227

Published online: 14 Sep 2012.

Applied Artificial Intelligence, 26:787–807, 2012
Copyright © 2012 Taylor & Francis Group, LLC
ISSN: 0883-9514 print/1087-6545 online
DOI: 10.1080/08839514.2012.718227

DESIGNING EFFICIENT ANN CLASSIFIERS FOR MATCHING BURGLARIES FROM DWELLING HOUSES

Mohammad Reza Keyvanpour[1], **Mohammad Reza Ebrahimi**[2], **and Mostafa Javideh**[2]
[1]*Department of Computer Engineering, Alzahra University, Tehran, Iran*
[2]*Technology Incubation Center of Alzahra University, Tehran, Iran*

□ *Leveraging supervised learning methods is vital for predictive analysis of crime data, however, because of the complex dependencies of crime behavioral variables, classifying behavioral crime profiles is considered to be a demanding task. This paper presents two classifiers for matching single-offender crimes of the type: Burglary from Dwelling Houses (BDH). The first classifier, Multiclass MLP Crime Classifier (M^2C^2), leverages a multiclass topology to become capable of matching nonprolific offenders in addition to prolific offenders. This method will be useful for matching crimes to several local offenders in a particular district, and it is not suitable for classifying a large number of offenders. Contrarily, the second method, Ensemble Neural Network Crime Classifier (EN^2C^2), focuses on automating decision-making processes for crime matching through exploiting expert classifiers' outputs in a bagging ensemble approach. As demonstrated by evaluative experiments, M^2C^2 is an efficient approach for classifying small numbers of nonprolific and prolific offenders. The proposed method's performance was proved when compared with other common machine learning techniques.*

INTRODUCTION

The paper focuses mainly on utilizing classification techniques for single-offender Burglary from Dwelling Houses (BDH) crime matching. The BDH type was chosen because of its significant influence on society as an important volume crime. The research uses artificial neural network classifiers because of their acceptable classification accuracy and also their noise-tolerance capability in this domain. However, during the study, diverse machine learning classification methods were applied to the

This research has been carried out under grant no. 90-3-2275 from Alzahra University. The support of this university is appreciated.

Address correspondence to Mohammad Reza Keyvanpour, Department of Computer Engineering, Alzahra University, Vanak, Tehran, Iran. E-mail: keyvanpour@alzahra.ac.ir

domain and also were compared to each other. Utilizing data mining techniques (especially neural networks) for crime matching is considered a new area of research and there are a relatively few valid publications about this topic.

Related Work

Recently, intelligent investigation of burglaries, thefts, and robberies has been tackled from several aspects, such as case-based reasoning (Ribaux and Margot 1999), computer simulation (Furtado et al. 2009), artificial intelligence and machine learning (Oatley, Ewart, and Zeleznikow 2006), and time series forecasting (Deadman 2003; Cohen and Gorr 2005). Ribaux and Margot (1999) introduced a general inference framework for investigative crime analysis based on case-based reasoning applied to burglary crime data. They proposed five general inference structures, which cover different crime matching scenarios.

Crime forecasting techniques have also been applied to the burglary crime domain. Fundamentals of crime forecasting can be found in a technical report from a practical project with a U.S. National Institute of Justice's grant (Cohen and Gorr 2005). Techniques for residential burglary crime forecasting are discussed in Deadman (2003) and Liu and Brown (2003).

Furtado and colleagues have proposed a multiagent-based simulation approach to understanding the behavior of criminals involved in crimes against property by using ant colony optimization and genetic algorithm (Furtado et al. 2009). They have presented a model for simulating real-world entities, including offenders, police patrol teams, and targets (public places). They have also simulated a criminal's learning ability.

A general BDH crime matching framework was proposed in (Keyvanpour, Javideh, and Ebrahimi 2011), which utilized data mining techniques to cover different crime matching scenarios. The OVER project (Oatley et al. 2006) and also Adderley's practical PhD thesis (Adderley 2007) may be considered to be the most prominent examples of using machine learning and artificial intelligence techniques in the BDH domain over the past decade. Oatley et al., the researchers in the OVER project, have assessed AI techniques including logic programming and case-based reasoning using similarity metrics such as *k*-nearest neighbor, Cosine Rule, and Tversky's Contrast model for matching single BDH crimes. As the authors in the OVER project have mentioned, KNN and Cosine Rule are not suitable for expressing similarities among crime behavioral patterns because they are geometric representations of similarity. The authors confirmed that none of the proposed mentioned approaches are robust predictive methods for crime matching.

Finally, in 2007, Adderley applied supervised and unsupervised neural networks for offender profiling in his PhD thesis (Adderley 2007). He used the multi-layer perceptron (MLP) for matching BDH instances with single prolific offenders and the self-organizing map (SOM) architecture for identifying distraction burglaries and sexual offences.

Contribution of the Article

Apart from Adderley's method, no robust and accurate supervised neural network model has been proposed for BDH crime matching. In this article, we provide a practical, intelligent investigation approach to burglary crime matching from a hybrid view of machine learning and data mining perspective. This approach focuses on different ways of using supervised neural network for BDH crime matching. Different types of classifiers were tested against a real BDH dataset and the results were evaluated. We also aim to improve the supervised method proposed in Adderley's thesis and eliminate its drawbacks through introducing two models named Multiclass MLP Crime Classifier (M^2C^2) and Ensemble Neural Network Crime Classifier (EN^2C^2). The former is proposed to eliminate a restriction in Adderley's structure that it is suitable only for prolific offenders, and the latter aims to provide a decision-making mechanism on identifying the most deserving offenders by leveraging an ensemble learning approach, an important feature that Adderley's method lacks. We used neural networks in both of these models because of their outstanding abilities with regard to noise-tolerance and desirable classification accuracy (Dreyfus 2005).

The organization of this article has been inspired by the main processes of the cross industry standard process for data mining (CRISP-DM) methodology (Shearer 2000), which is widely used in data mining. Consequently, different aspects of business understanding are described in the following section. Data understanding and data preparation are proposed in "Data Understanding and Preparation." Model building and proposed methods are dissected in "Proposed Classifier Models." Evaluation issues are discussed in the "Experimental Results."

BUSINESS UNDERSTANDING

The goal of crime matching in law enforcement agencies is to assign the proper crime to the proper offender and/or vice versa. As Burgess et al. (2007) stated, "The first possible benefit of a classification system would be in aiding the apprehension of the offender through behavioral investigative profiling. "During this work, the authors studied the business process of solving BDH crimes currently used by investigators in the capital city of

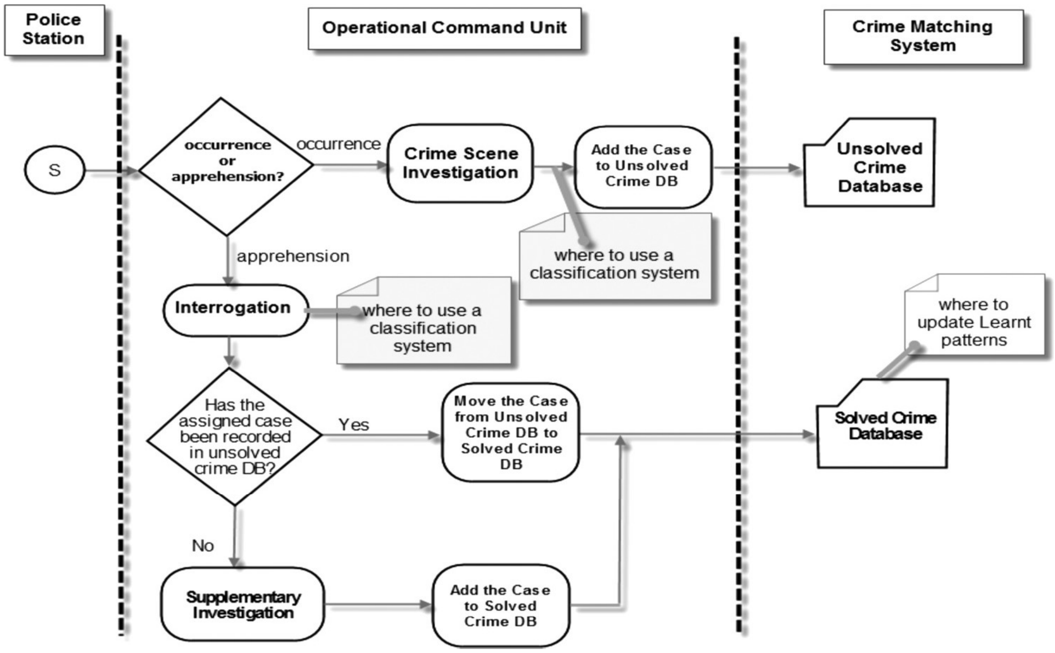

Police Station
Operational Command Unit
Crime Matching System
S
occurrence or apprehension?
occurrence
Crime Scene Investigation
Add the Case to Unsolved Crime DB
Unsolved Crime Database
where to use a classification system
apprehension
Interrogation
where to use a classification system
where to update Learnt patterns
Has the assigned case been recorded in unsolved crime DB?
Yes
Move the Case from Unsolved Crime DB to Solved Crime DB
Solved Crime Database
No
Supplementary Investigation
Add the Case to Solved Crime DB

FIGURE 1 Intelligent crime matching schema.

Iran. Subsequently, an intelligent version of this process was proposed that embeds BDH classifier systems in two stages of the process.

The proposed business flow of the crime matching process is demonstrated in Figure 1. Dark callout symbols illustrate the stages where a BDH classifier can be utilized in this process. As can be seen in the figure, the process is started when (1) a new unsolved burglary instance is reported (occurrence mode), or (2) a specific burglar is apprehended (apprehension mode). In the first mode, a classifier helps to assign several probable recorded offenders to the burglary instance, and in the second mode the classifier participates to estimate the most probable instances that the burglar could have done. Through the following sections, we will describe how supervised neural network classifiers can be used for aiding the two above-mentioned tasks. Adderley's has covered a special case of the second mode by his proposed MLP neural network (the model is special because it can handle only prolific offenders). We first propose a supervised method named M^2C^2 for eliminating a restriction of his method. Then we propose another supervised method (EN^2C^2) to cover the first mode.

DATA UNDERSTANDING AND PREPARATION

There were 160 real BDH cases involved in this work. Real cases of BDH crime were collected from two Operating Command Units (OCUs) in Iran.

The data were extracted from real textual police narrative reports and transformed into a binary modus operandi (MO) format (see Keyvanpour, Javideh, and Ebrahimi [2011] for a detailed discussion on the binary encoding process of BDH crimes). Finally, extracted crime features were manipulated through consulting two BDH experts.

As another data preparation activity, all of the categorical data types were transformed into binominal fields using binary set encoding. This kind of transformation is a common preprocessing task for building MOtables representing offenders’ behaviors (Adderley 2007; Oatley, Ewart,

TABLE 1 Structure of the Sample Data

Case type	Offender no.	Number of burglaries	Dominant behavioral pattern
Ordinary Cases (majority)	of #1	22	EntryLocation: “terrace”; EntryMethod: “smash”; Building Type: “apartment”; Instrument: “lever”; StolenProps: “small-jewelry”
	of #2	5	Building Type: “bungalow”; ExitLocation: “same as entry”; ReagonSpec: “rich”
	of #3	28	EntryLocation: “window”; EntryMethod: “cut”; Building Type: “bungalow”; ExitLocation: “same as entry”; StolenProps: “small- jewelry”
	of #4	26	EntryLocation: “Main door”; EntryMethod: “smash”; Building Type: “apartment”; Instrument: “lever”;ExitLocation: “same as entry”
	of #5	3	Building Type: “apartment”; SearchType: “untidy”; StolenProps: “only jewelry”; ReagonSpec: “rich”; Instrument: “lever”
	of #6	6	EntryLocation: “terrace”; EntryMethod: “neighbor houses”; Building Type: “apartment”; SearchType: “untidy”; ReagonSpec: “rich”
	of #7	15	EntryLocation: “terrace”; OccurrenceTime: “around midnight”; Building Type: “apartment”; ReagonSpec: “rich”
	of #8	20	SearchType: “tidy”; OccurrenceTime: “before noon”; Building Type: “apartment”; StolenProps: “big”; ReagonSpec: “rich”
	of #9	5	EntryLocation: “wall”; Building Type: “flat”; StolenProps: “small- jewelry”
	of #10	14	EntryLocation: “wall”; Building Type: “apartment”; OccurrenceDay: “weekend”; OccurrenceTime: “around midnight”; StolenProps: “small-jewelry”; ReagonSpec: ” rich”
Challenging Cases	of #11	5	Building Type: “bungalow”; ReagonSpec: “rich”
	of #12	6	EntryMethod: “Bogus”; Building Type: “apartment”; StolenProps: “only jewelry”
	of #13	3	Building Type: “bungalow”; StolenProps: “only jewelry”; ReagonSpec: “rich”
	of #14	2	EntryLocation: “wall”; OccurrenceTime: “around midnight”
Total	14	160	–

and Zeleznikow 2006). This transformation was done because there are rather few classifiers capable of directly handling categorical data in the literature for machine learning and data mining compared with that existing for dealing with non categorical data. Also, it was admitted that investigators feel more comfortable with yes/no questions when making their related reports. 187 binary features were constructed this way. The encoded binary data contained information about a burglar's behavior at the crime scene, his or her method of committing the burglary, victimized building specifications, and so on.

In the initial feature selection step, all of the features with variance 0 were removed from the dataset. Also, less-important features were identified and omitted through consulting a domain expert. Finally, 94 features remained to be involved in the model-building phase. In order to examine the noise-tolerance capability of the model, 16 noisy cases were added to the dataset, mentioned as challenging cases in Table 1.

These cases include three types of data: (1) Nonprolific offenders who had a small number of ordinary offences; in fact, they had not committed more than six instances and their crime instances had not included any particular MO pattern that could be distinguished from the other offenders' patterns. (2) Some intentionally added noisy data; these noisy data were designed to mimic the data entry errors a human agent makes when entering the crime behavioral information into the system. (3) Crime instances including a high rate of missing values; the rationale behind adding challenging cases was to urge the model to face with instances that are not easy to classify. As can be seen in Table 1, there are 160 BDH instances committed by 14 single offenders, each one has committed from 2 to 28 cases. The last column in the table contains the main crime patterns for each offender. These patterns were estimated through consulting a domain expert to provide a preliminary insight into data.

CASTING BDH CRIME MATCHING PROBLEM TO SUPERVISED ANN CLASSIFICATION: ADDERLEY'S METHOD

The general crime matching problem was discussed in the "Business Understanding" section. The crime matching problem is equivalent to a classification or clustering problem from the machine learning point of view. Because we aim to address supervised learning methods in this article, we will focus on the classification problem. Thorough overviews on the BDH crime matching concept have been presented in (Ribaux and Margot 1999). Presently, the naive supervised MLP proposed by Adderley is considered the latest approach of using ANNs for supervised crime matching. Accordingly, in this section, we deliberate on Adderley's method of using neural networks for burglary crime matching. Different proposed classifiers' architectures are dissected in the following sections.

Dissecting the Method

Adderley proposed an elegant approach for assigning new unsolved BDH instances to prolific offenders in his PhD thesis. He used 22 binary-encoded MO variables (Keyvanpour, Javideh, and Ebrahimi 2011) in the network's input layer and a single neuron representing the network's yes or no answer as the output layer. As it can be seen in Figure 2, using this neural network topology, the input layer should have as many neurons as extracted MO features and the output layer should have just one neuron. Also, the real output value should be interpreted as a binary value. Using this model, when a new unsolved crime occurs, it should be passed to each of the trained prolific offender's network to be assessed with regard towhether it can be attributed to that offender. This model suffers from the three following drawbacks:

1. In Adderley's model, in order to obtain at least 50% of accuracy, each offender's network should be provided with at least 35 solved BDH training instances of a particular offender. Because the approach needs a significant number of training instances (all related to a particular burglar) to be trained, it will apply only to prolific offenders. Foley's rule (Priddy and Keller 2005) implies that if the ratio of inequality (1) is satisfied, the resultant error of the classifier will be close to that of Bayes optimal classifier.

$$\frac{n(S)}{n(F)} > 3, \tag{1}$$

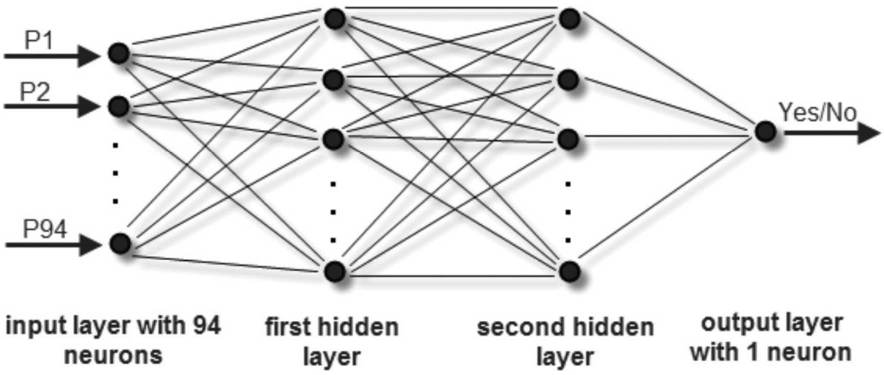


FIGURE 2 Trained MLP neural network for a prolific burglar.

where $n(\mathrm{S})$ is the number of training samples per class and $n(\mathrm{F})$ is the number of features (22 in Adderley's approach). So, the optimal size training set required for $n(\mathrm{Of})$ offenders in this model can be calculated by the following formula:

$$n(\mathrm{S}_{train}) > 3n(F) \times n(C) \times n(Of), \tag{2}$$

where $n(\mathrm{C})$ is the number of classes (one yes class and one no class in this model). It means that, in order to have an ideal classifier for a specific offender, the model should be provided with at least $3 \times 22 \times 2$ solved crimes attributed to a specific offender in the training phase, which requires the offender to be prolific. The general neural network architecture presented in "Dataset Issues" reduces the number of optimal training sets to alleviate this deficiency.

2. This model cannot be generalized for use in the second mode (see "Business Understanding"). No systematic approach has been proposed to help decide which offender's network output is more reliable than that of the others. Generating only a list of confidence values produced by each offender's network is not satisfactory. For example, a situation corresponding networks for offender X and offender Y produces confidence values 0.71 and 0.74, respectively; simply choosing the offender Y because of its maximum confidence value does not seem a wise decision. In such a situation, both of the networks are not so confident and it is also possible that the offender X might be the correct offender. Because the decision may affect someone's freedom, it is worthwhile to exploit an ensemble mechanism to enrich the quality of decision making. An ensemble approach to alleviate this deficiency has been proposed in the next section.
3. It seems that using just 22 crime variables as input features is not sufficient to provide the best discriminative power. This has been proved in the practical OVER project, which has utilized over 120 crime features (Oatley, Ewart, and Zeleznikow 2006). We extracted 94 input features to eliminate this drawback. This claim will be proved in the evaluation of the experiments in "Experimental Results."

We applied the architecture of MLP neural network proposed by Adderley (Adderley 2007) against our work's dataset for BDH supervised crime matching (Figure 2). We used the following step function as the activity function of the last neuron to perform a simple Boolean interpretation:

$$\mathrm{f}(\mathrm{net}^{\mathrm{output-layer}}) = \begin{cases} 0 & \mathrm{net}^{\mathrm{output-layer}} < \mu \\ 1 & \mathrm{net}^{\mathrm{output-layer}} \geq \mu \end{cases}, \tag{3}$$

where μ is a threshold parameter with a positive value less than 1. The output layer's input is $\text{net}^{\text{output-layer}}$, which comes from the last hidden layer and also is scaled between 0 and 1 using a min-max normalization function (Equation (4)). It is important that this value can be interpreted as a measure for confidence. If the value is near 0.5, it can be induced that we cannot rely on the network's answer. (e.g., if the output layer produces value 0.55). We use this value as the confidence measure of classifier response and name it CV for the remainder of this paper. CV can be calculated by the following equation:

$$CV = \text{net}^{\text{output}-\text{layer}} = \text{Min\_Max}\left(\sum_{i=1}^{\text{n}} w_i \times \text{net}_i^{\text{last}-\text{hidden}-\text{layer}}\right), \tag{4}$$

where n is the number of neurons resident in last hidden layer, Wi is the corresponding weight between ith neuron of the last hidden layer and the output layer. The net input of the ith neuron in the last hidden layer is $\text{net}_{\text{i}}^{\text{last}-\text{hidden}-\text{layer}}$.

It is worthwhile to note that tuning the threshold parameter μ can affect the false positive and false negative rate of the classifier. Roughly stated, increasing this value may result in a lower false positive rate and vice versa. Presenting an optimal value for this parameter would require another major study, so we will leave the issue by simply using $\mu = 0.5$ in our work. The remainder of this paper has been dedicated to discussing the proposed methods of BDH crime classification and also to presenting the experimental results.

PROPOSED CLASSIFIER MODELS

To alleviate the two drawbacks mentioned in "Dissecting the Method", two different crime classifiers are introduced in this section. The first classifier exploits multiclass neural network architecture to cover nonprolific offenders in addition to prolific offenders. The method aims to reduce the number of required training samples by half. This way, the classifier might be capable of matching both prolific and nonprolific offenders with their related crime instances. For the sake of simplicity, we refer to this method as M^2C^2 (Multiclass MLP Crime Classifier). It is important to note that using a multiclass topology may increase the model's training time (Priddy and Keller 2005), but it will be useful when we are dealing with several nonprolific burglars who are primarily active in a specific geographic area. We will dissect the classification accuracy issues in "Experimental Results."

The second proposed classifier, which is preferred by the authors, is an ensemble classifier addressed as EN^2C^2 (Ensemble Neural Network Crime Classifier) through the remainder of the article. EN^2C^2 aims to present a general remedy for the second drawback mentioned previously. That is, it will present a decision regarding which offender is preferred to be attributed to a crime instance when there are more than one offender with almost the same confidence value (CV).

M^2C^2: Multiclass MLP Crime Classifier

A practical approach for learning the patterns among solved crime instances is to design the ANN topology to have as many neurons as the total number of offenders in the output layer. We used a softmax activity function in the output layer of this topology:

$$f(\mathrm{net}_i) = \frac{\mathrm{e}_i^{\mathrm{net}}}{\sum\limits_{j=1}^{n=14} \mathrm{e}_j^{\mathrm{net}}}, \tag{5}$$

where, net_i (net_j) is the last hidden layer output for ith (jth) output neuron, $n = 14$ is the number of output neurons (number of offenders in our dataset) in this topology. Utilizing softmax function as the output layer's activity function ensures that all of the output values will be between 0 and 1, and also their sum will be 1. This way, we were able to interpret the output values and also rank the predicted offenders according to their corresponding confidence values. Accordingly, the confidence value for assigning crime instance $\mathrm{P} = (p_1, \ p_2, \ldots, p_{94})$ to ith offender is given by Equation(6).

$$CV_i = \mathrm{softmax}(\mathrm{net}_i^{\mathrm{output-layer}}). \tag{6}$$

Figure 3 illustrates our general multiclass topology for classifying solved crime instances.

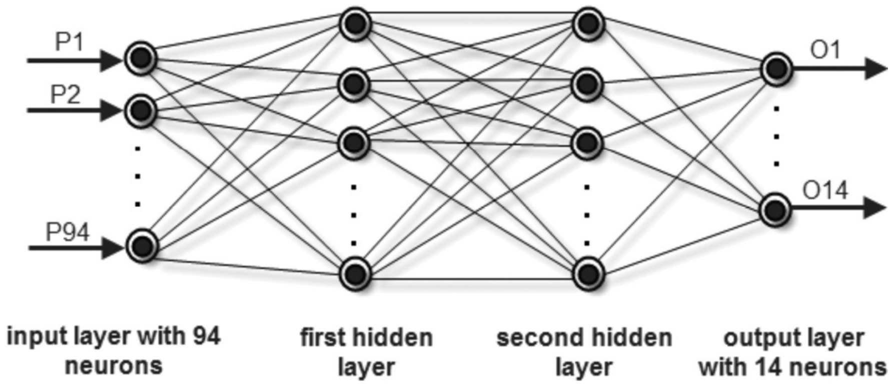


FIGURE 3 Multiclass network topology for crime matching.

In comparison with the topology that was discussed in "Data Understanding and Preparation," in this method the number of classes (n(c)) has been reduced by half, so according to inequality (2), the model can be trained optimally using smaller training datasets. As already mentioned, although in comparison with previous topology, this approach may result in increase of training time; it is capable of covering nonprolific local offenders who usually commit most of their crimes in particular areas. So it can be exploited by the OCUs for matching burglaries occurred in a specific district. This way, the network finds the opportunity to learn the behavioral patterns of nonprolific offenders in comparison with that of other offenders. Thus, it is important to note that M^{2}C^2 is not applicable for classifying large numbers of offenders because its topology does not allow having too many neurons in the output layer. EN^{2}C^2,which will be proposed in the next section, can also eliminate this deficiency.

EN^{2}C^2: Ensemble Neural Network Crime Classifiers

This section is dedicated to presenting an ensemble learning approach for eliminating the deficiency of lacking a decision-making mechanism for preferring a reliable offender's network over other reliable networks. The key idea to using an ensemble approach in this domain is the fact that exploiting multiple highly trained classifiers that have learned the style (pattern) of one offender can perform better than a single classifier that has partially learned the patterns of all offenders. Figure 4 depicts the

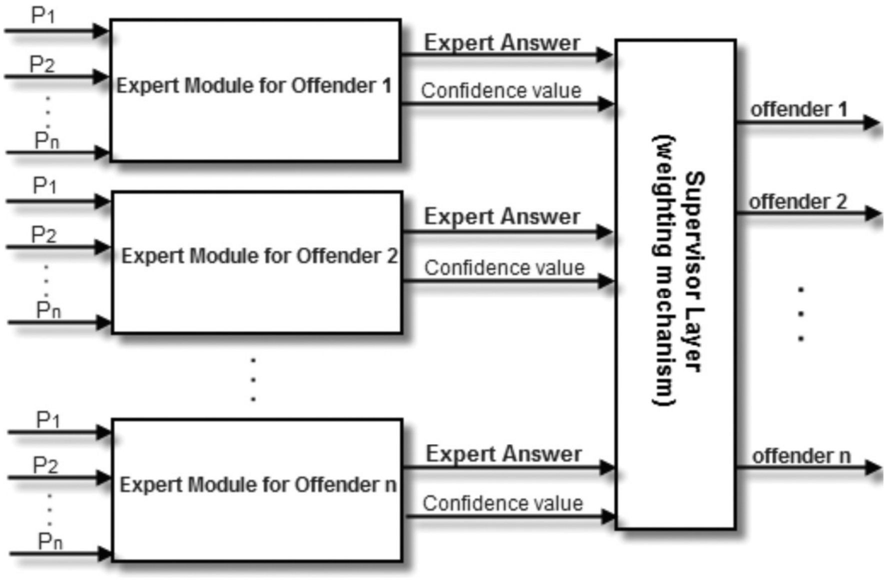


FIGURE 4 General schema of a modular crime classifier.

general schema for a modular crime classifier. As the figure shows, the entire classifier is composed of several expert classifiers, each of them trained for a specific offender. Each expert delivers its output to a supervisor layer, which is responsible for weighing the outputs of experts according to their delivered confidence values. This layer may also be responsible for selecting the more confident classifier and identifying it as the final output of the ensemble. EN^{2}C^2 generally utilizes bootstrap aggregation, commonly referred as the bagging (Duda, Hart, and Stork 2001) technique of machine learning. As the name suggests, the model assembles several binary crime classifiers, which are known as experts. In fact, each binary classifier is expert in identifying its own offender's pattern. This approach can work well if the component classifiers are experts in separate regions of the input space (Duda, Hart, and Stork 2001).

The main architecture of EN^{2}C^2 (Figure 5) consists of two layers: (1) the expert layer and (2) the supervisor layer. The first layer contains binary classifiers of each offender and the second is a multiclass MLP network. Upon their classification accuracy, MLP neural networks were chosen as the expert classifiers of EN^{2}C^2 (see "Experimental Results"). Training the first layer can be accomplished in parallel whereas training the second layer must be done after the first. As can be seen in Figure 5, the binary output of each expert (yes/no) in addition to expert's level of confidence (CV) will be delivered to a supervisor MLP neural network. This supervisor network is trained to identify the correct offender based on received information from expert classifiers. In fact, the supervisor network weighs each binary classifier according to its corresponding binary answer and also the related

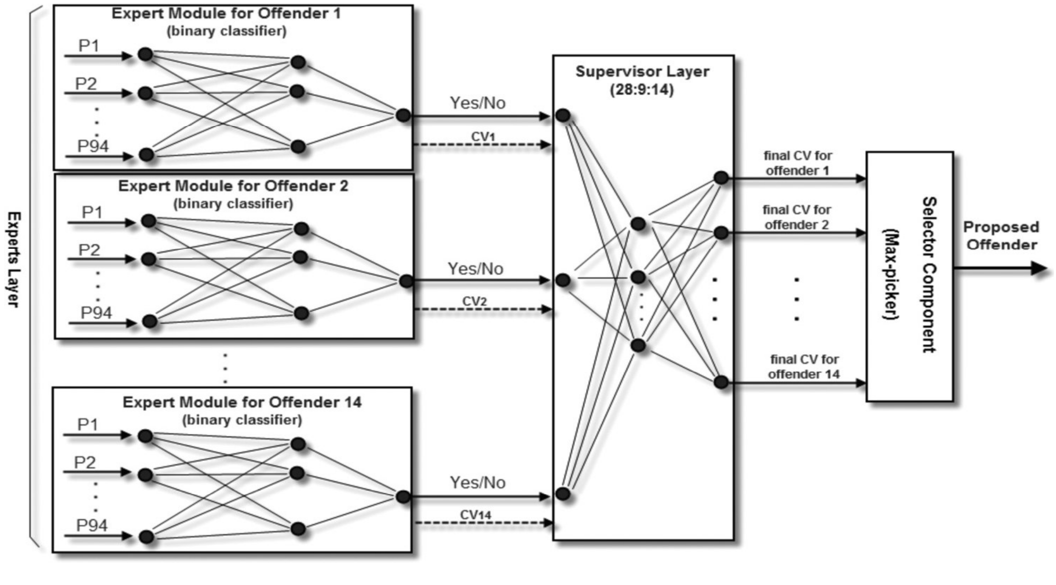


FIGURE 5 The architecture of EN^{2}C^2.

value of CV (see Equation (4) for calculation of CV). Eventually, according to the supervisor network's output, a selector component (which is simply a max-picker) decides on the offender to propose for the input crime instance. The supervisor network must have 14×2 neurons in the input layer, 14 neurons (number of offenders in the dataset) in the output layer. After testing the accuracy of different network structures, finally we designed this network with one hidden layer containing nine neurons.

In the EN^2C^2 framework, all of the activity functions used in non-output layers in both expert classifiers and supervisor layer were the popular sigmoid nonlinear function, which is given by the following equation for the *i*th unit in a layer:

$$\mathrm{f}(\mathrm{net}_i) = \frac{1}{1 + \mathrm{e}^{-\mathrm{net}_i}}. \tag{7}$$

For the output layers in expert classifiers, step function of Equation (3) was used and finally, for the output layer of the supervisor network, softmax function of Equation (5) was used. It is important to note that, using the softmax function ensures the output values of the supervisor layer lie between 0 and 1. So the Equation (8), which is considered a principle in the bagging technique always holds:

$$\sum_{i=1}^{c} W(CVi) = 1, \tag{8}$$

where *c* is the number of expert classifiers, which is equal to the number of offenders, and W(CV*i*) is the output of *i*th neuron in the output layer of the supervisor layer. In the other words, $\mathrm{W}(\mathrm{CV}_i)$ is considered to be the modified confidence value for *i*th expert classifier. This value has been represented as final CV for offender *i*' in Figure 5.

EXPERIMENTAL RESULTS

In this section we have evaluated the performance of M^2C^2 and EN^2C^2. Also, we have compared several common classifiers that have been applied on the problem domain.

Dataset Issues

Unfortunately, for security issues, there is not a standard common dataset for the crime matching domain. Consequently, it is almost impossible for the researchers across different countries to access each others' datasets. So most of the researchers usually leverage a systematic method of data collection and build their own real datasets with the help of their countries' law enforcement agencies. The structure of the sample data set used in this study was discussed in "Data Understanding and Preparation." We intentionally collected the dataset records to include both prolific and nonprolific offenders. Thus, about 36% of the records are related to offenders with fewer than 5 crimes. As did Adderley and Oatley, Ewart, and Zeleznikow, we used a systematic approach for collecting required data from real BDH cases of two OCUs and, subsequently, we built our work's dataset after performing an initial preprocessing task. As mentioned, for the evaluation purposes we designed the dataset to include offences of both prolific and nonprolific burglars.

Evaluation Method

Because a BDH classifier can affect a person's freedom, classification accuracy has been used as the most important performance measure of BDH crime classification in (Adderley, 2007; Oatley, Ewart, and Zeleznikow 2006). Training time and robustness are other classification performance measures. Because designing classifiers in the domain of crime matching usually considers peoples' freedom, training time does not play a decisive role and even a few days of training might be acceptable in this domain. Contrarily, robustness of the classifier is an important issue, because the system might be used with different numbers of crimes representing diverse distributions of BDH incidents in each OCU's district.

We assessed the classification accuracy, as our performance measure, by leveraging a 5-fold cross-validation technique (see Kohavi [1995]). Using this technique helps us to evaluate the robustness of the classifier and also to avoid network overfitting. Each fold contained 32 systematically sampled crime instances with no overlap.

Experiment 1: Evaluating M^2C^2

In order to evaluate the performance of M^2C^2, we assessed the classification accuracy of the model separately in each cross-validation fold. The classification accuracy of each fold in addition to misclassified instances is shown in Table 2. The average of classification accuracy is about 92.5% in this method. We also assessed the confidence value of the classifier for each fold. Figure 6 depicts the level of confidence (vertical axis) for each crime instance (horizontal axis) in each fold. As already mentioned, confidence value is a real number between 0 and 1. As the figure shows, all of the five folds have at least the confidence value of 90%.

TABLE 2 Classification Accuracy of M^2C^2

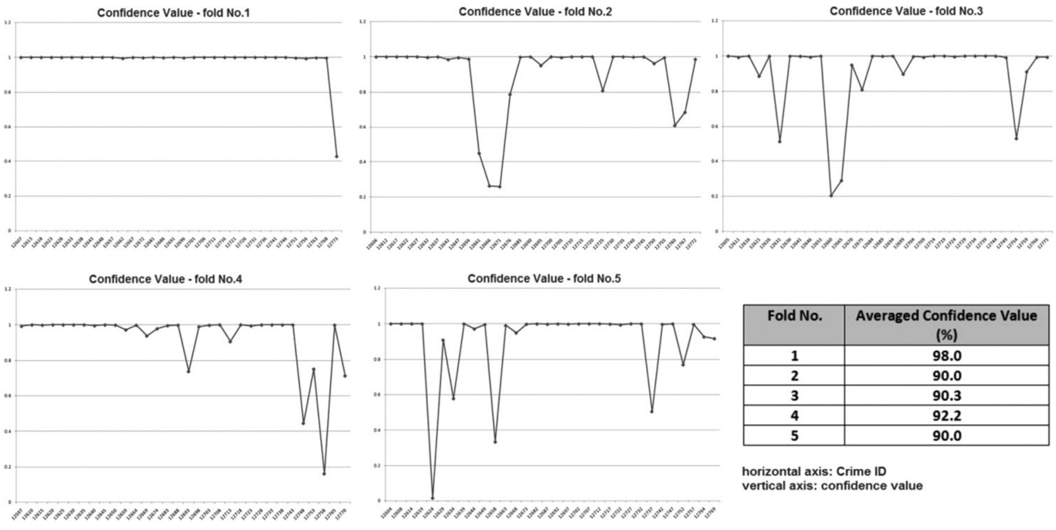
Fold No.	Accuracy (%)	ID of misclassified instances
1	93.7	12768,12773
2	93.7	12772,12671
3	96.9	12759
4	90.6	12770,12753,12748
5	87.5	12752,12737,12658,12634

It can be also observed that fold number 5 has the lowest rate of the confidence level compared with the other cross-validation folds. This fold also has the lowest classification accuracy as seen in Table 2. This reveals the fact that when the network is not confident enough, it may not provide good classification accuracy. In contrast, the smooth shape of fold number1 can be attributed to its high percentage of the instances that are related to prolific offenders with distinct behavioral patterns. So it is relatively easy for the classifier to accurately identify these crime instances (accuracy of 93.7% and confidence value of 98%).

Results of the experiment revealed that although it is also hard for M^2C^2 to classify nonprolific offenders, it presents acceptable outcomes (at least 87.5% accuracy and 90% confidence value).

Experiment 2: Evaluating EN^2C^2

We evaluated the classification accuracy of EN^2C^2 by the same five folds that we used for evaluating M^2C^2 through the 5-fold cross-validation



Fold No.	Averaged Confidence Value (%)
1	98.0
2	90.0
3	90.3
4	92.2
5	90.0

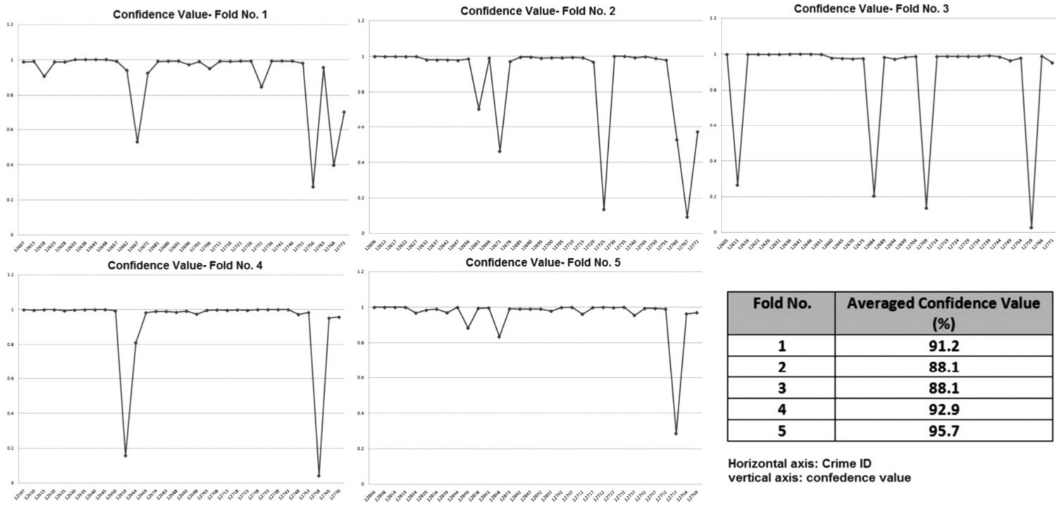
FIGURE 6 Confidence analysis of M^2C^2 (5-folds cross-validation).

TABLE 3 Classification Accuracy of EN^{2}C^2

Fold no.	Accuracy (%)	ID of misclassified instances
1	93.7	12672,12773
2	81.2	12671,12725,12750,12760,12767,12772
3	84.3	12771,12759,12709,12684,12665
4	96.9	12659
5	100	–

process. The accuracy related to each fold and also misclassified instances has been shown in Table 3. Surprisingly, the experiment reports the averaged classification accuracy of 91.2% with this method, which is a value near to the accuracy of M^{2}C^2 method. Regarding the relatively low training quality of expert modules of nonprolific offenders, it was expected that EN^{2}C^2 would provide a lower accuracy than M^{2}C^2. However, it turned out that EN^{2}C^2 may also perform acceptably for even nonprolific offenders, and its overall performance is close to that of M^{2}C^2.

By comparing Tables 2 and 3, it can be inferred that M^{2}C^2 has performed better in folds 2 and 3, but EN^{2}C^2 has performed better in folds 4 and 5. Also, both approaches have performed equally in fold 1. The results of confidence analysis are shown in Figure 7. According to the figure, folds 2 and 3 have the lowest confidence level and they have the lowest level of accuracy, too (Table 3). As we saw for M^{2}C^2, the evaluation results again endorse the hypothesis that the network will not provide good classification accuracy if it is not confident enough. It was also admitted that the average of confidence levels of EN^{2}C^2 (91.2%) was also near to that of M^{2}C^2 (92.1%).



Fold No.	Averaged Confidence Value (%)
1	91.2
2	88.1
3	88.1
4	92.9
5	95.7

FIGURE 7 Confidence analysis of EN^{2}C^2 (5-folds cross-validation).

Experiment 3: Comparative Evaluation

As a comparative evaluation, we have explored the accuracy of the proposed methods of crime classification in comparison with common methods recently applied by researchers on crime matching and prediction domain. These common methods include MLP networks (Adderley 2007), radial basis function networks (RBFNs) (Lv, Ji, and Zhang 2008), inductive algorithms such as decision trees (C5.0 and CART; Oatley, Ewart, and Zeleznikow 2006), Bayesian belief nets (Oatley, Ewart, and Zeleznikow 2006), and support vector machine (SVM). Their level of accuracy has been assessed in a comparative approach (Figure 8).

The bar chart in Figure 8 illustrates the accuracy of first to fifth cross-validation fold, respectively from left to right, in addition to the averaged accuracy for each algorithm. The chart reveals that both M^2C^2 and EN^2C^2 provide the highest rates of accuracy with 92.5% and 91.2%, respectively. This shows that the proposed models perform accurately in comparison to the other common machine learning methods. As the results show, C5.0 and RBF network are the second most accurate (87.5%). Also, it can be claimed that SVMs cannot provide acceptable accuracy in this domain of application.

In this part of the article, it's worthwhile to mention a fact about Bayesian network classifiers, which we encountered in evaluative experiments. Although Bayesian networks did not offer the best accuracy in this study, they bring a significant benefit into the crime matching domain that other powerful classifiers do not. The benefit of this method has been illustrated in Figure 9. In this case, for the purpose of illustration, three offenders (#4, #7, and #9) were selected with 26, 15, and 5 attributed crimes,

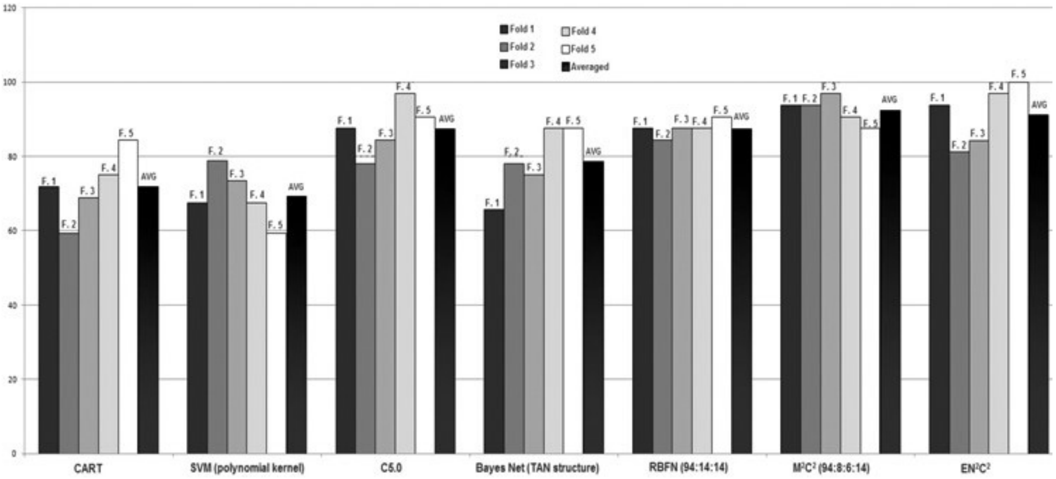


FIGURE 8 Confidence analysis of EN^2C^2 (5-folds cross-validation). (Color figure available online.)

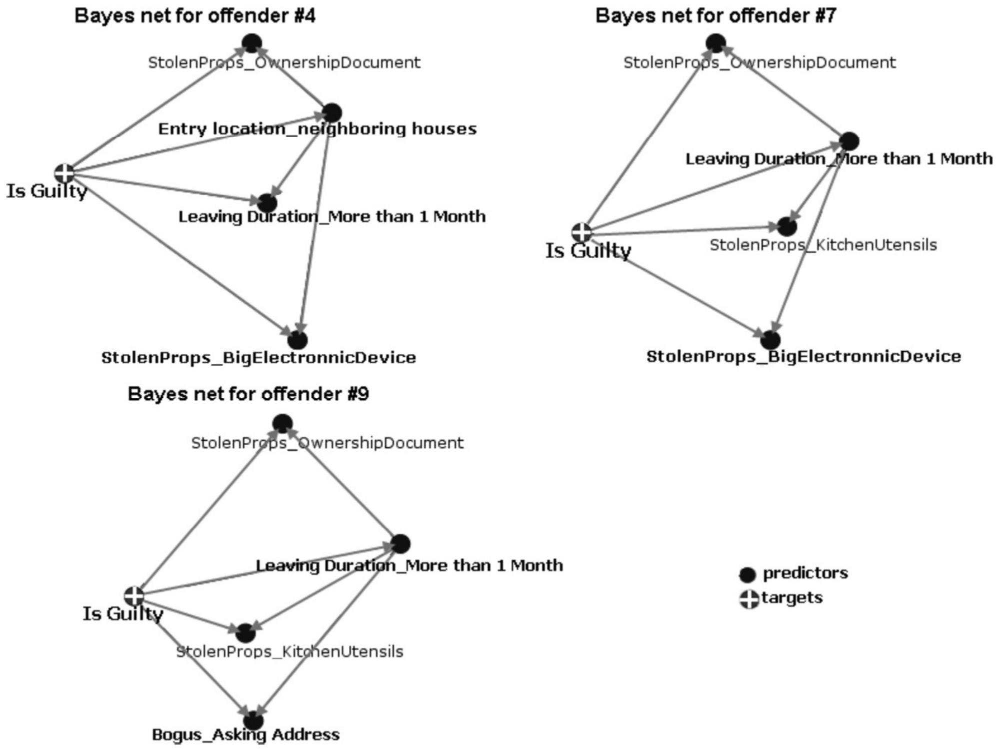

Bayes net for offender #4
StolenProps_OwnershipDocument
Entry location_neighboring houses
Is Guilty
Leaving Duration_More than 1 Month
StolenProps_BigElectronnicDevice
Bayes net for offender #7
StolenProps_OwnershipDocument
Leaving Duration_More than 1 Month
Is Guilty
StolenProps_KitchenUtensils
StolenProps_BigElectronnicDevice
Bayes net for offender #9
StolenProps_OwnershipDocument
Leaving Duration_More than 1 Month
Is Guilty
StolenProps_KitchenUtensils
Bogus_Asking Address
● predictors
targets

FIGURE 9 TAN Bayesian networks trained to predict the guiltiness of a specific offender.

respectively. Offender #4 is considered to be an instance of prolific burglars and offender #9 may be considered a nonprolific offender. As the figure depicts, this kind of classifier is considered as a graphical tool that is able to illustrate the behavioral signature of the criminal in a comprehensible manner. This graphical model may provide a deep insight into the data for police investigators. We used a tree-augmented naïve Bayes (TAN) structure (see Friedman, Geiger, and Goldszmidt [1997]) which preserves the dependencies among predictors as a tree structure.

After feature selection, four predictors (i.e., independent crime variables) were used as the Bayesian network's inputs.

Comparative Evaluation of Experts Layer in EN^{2}C^2

In order to evaluate the performance of expert modules, which we used in the expert layer of EN^{2}C^2, we compared different common models of binary classifiers. We applied diverse binary classifiers against all of the offenders in the dataset. As noted earlier, EN^{2}C^2 uses MLP networks as its component classifiers in the experts layer (see Figure 5).

Table 4 shows the experiment's results for different structures of these component classifiers. As can be seen in the table, each binary classifier has

TABLE 4 Comparing Best Performance of Different Binary Classifiers for Use in Experts Layer

	Classification accuracy(%)													
Classifier type	Offender #1	Offender #2	Offender #3	Offender #4	Offender #5	Offender #6	Offender #7	Offender #8	Offender #9	Offender #10	Offender #11	Offender #12	Offender #13	Offender #14
MLP (1hidden layer)	98.9 (94:6:1)	100 (94:7:1)	100 (94:7:1)	98.8 (94: 8:1)	100 (94:6:1)	100 (94:7:1)	98.7 (94:7:1)	100 (94:6:1)	100 (94:10:1)	100 (94:6:1)	100 (94:6:1)	100 (94:6:1)	98.5 (94:7:1)	97.3 (94:10:1)
MLP (2hidden layers)	100 (94:4:5:1)	100 (94:2:2:1)	100 (94:6:3:1)	100 (94:2:3:1)	100 (94:2:5:1)	100 (94:2:2:1)	100 (94:3:5:1)	100 (94:3:4:1)	100 (94:4:8:1)	100 (94:3:5:1)	100 (94:3:4:1)	100 (94:2:2:1)	100 (94:2:2:1)	100 (94:4:5:1)
RBFN	96.5 (94:14:1)	97.6 (94:14:1)	100 (94:14:1)	100 (94:14:1)	100 (94:14:1)	97.6 (94:14:1)	98.6 (94:14:1)	100 (94:14:1)	100 (94:14:1)	98.7 (94:14:1)	97.7 (94:14:1)	100 (94:14:1)	100 (94:14:1)	98.7 (94:14:1)
C5.0	97.6	98.8	100	98.8	100	100	98.8	100	100	98.8	100	98.8	98.8	100
CART	97.6	95.2	97.6	97.6	100	100	95.2	97.6	96.4	94	96.4	98.8	97.6	98.8
Bayes net (4 predictors)	79.52	85.6	77.1	77.1	87.9	89.2	82	74.7	89.2	78.3	85.5	85.5	87.9	90.4
SVM	91.6	96.4	92.8	98.8	98.8	96.4	89.2	90.4	96.4	91.6	96.4	95.2	98.8	98.8
Logistic Regression	53	35	36.1	77.1	71	84.3	53	79.6	81.9	61.4	82	84.3	83.1	77.1

been able to deal with its own related offender accurately. RBF networks and 2-hidden-layer MLPs outperformed the other binary classifiers. Additionally, it can be seen that 2-hidden-layer MLPs are able to predict a specific offender's crimes accurately using fewer numbers of neurons than RBFNs. To be provided with a high rate of accuracy is the main essence of using an expert classifier in an ensemble approach and justifies choosing ANN classifiers in the expert layer of EN^2C^2. Surprisingly, the results of the experiment in Table 3 reveal that EN^2C^2 is also capable of classifying nonprolific offenders' crimes although we developed the method with the main motivation of utilizing ensemble learning in a crime matching process.

CONCLUSION

Customized machine learning classifiers are efficient tools for crime matching tasks. Based on the results of this study, generally, ANN crime classifiers family (MLPs and also RBFNs) can result in better accuracy than other common classification methods such as C5.0, CART, Bayes net and SVMs, in the crime matching domain. Accordingly, we proposed two ANN-based classifiers, M^2C^2 and EN^2C^2, which aim to remove restrictions and drawbacks of existing ANN methods that applied in the domain. M^2C^2 was proposed to help investigators with the crime matching process when dealing with small numbers of prolific or nonprolific offenders. EN^2C^2 was proposed to support decision-making in crime matching process. Results of the study showed that M^2C^2 outperforms other types of classifiers, and it was capable of matching both prolific and nonprolific offenders with an acceptable level of accuracy (92.5%). This method will be useful for matching crimes to several local offenders who are primarily active in a special geographical district. Contrarily, EN^2C^2 focuses on supporting investigators with decision making through exploiting expert classifier outputs in a bagging ensemble approach.

Leveraging ensemble learning, EN^2C^2 also worked well on the dataset containing both prolific and nonprolific offenders and also presented the averaged accuracy of 91.2%. As a result, both methods might be leveraged by police investigators according to their requirements, however, the authors recommend using EN^2C^2 whenever it is possible. Also, experimental results showed that, although the popular Bayesian belief networks have not provided the best accuracy, they are very useful as a graphical tool, which can reveal the offender's crime behavior signature in a comprehensible format. So, the authors also recommend this kind of classifier to be used as a complementary tool for crime pattern visualization, because other types of classifiers are not capable of providing it.

REFERENCES

Adderley, R. W. 2007. "The Use of Data Mining Techniques in Crime Trend Analysis and Offender Profiling." PhD thesis, University of Wolverhampton, England.

Burgess, A. W., M. L. Commons, M. E. Safarik, R. R. Looper, and S. N. Ross. 2007. Sex offenders of the elderly: Classification by motive, typology and predictors of severity of crime. *Aggression and Violent Behavior* 12:582–597.

Cohen, J., and W. L. Gorr. 2005. *Development of crime forecasting and mapping systems for use by police* (Technical Report). John Heinz School of Public Policy and Management, Carnegie Mellon University, USA.

Deadman, D. 2003. Forecasting residential burglary. *International Journal of Forecasting* 19:567–578.

Dreyfus, G. 2005. *Neural networks: Methodology and applications.* Berlin, Germany: Springer.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern classification*, 2nd ed. New York: John Wiley Interscience.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. *Bayesian network classifiers.* Machine Learning 29. Netherlands: Kluwer Academic Publishers.

Furtado, V., A. Melo, A. Coelho, R. Menezes, and R. Perrone. 2009. A bio-inspired crime simulation model. *Decision Support Systems* 48:282–292.

Keyvanpour, M., M. Javideh, and M. Ebrahimi. 2011. Detecting and investigating crime by means of data mining: A general crime matching framework. In *Proceedings of Procedia Computer Science* 3:872–880.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on artificial intelligence* (IJCAI'95) 2. CA: Morgan Kaufmann.

Liu, H., and D. E. Brown. 2003. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting* 19:603–622.

Lv, L, N. Ji, and J. Zhang. 2008. A RBF neural network model for anti-money laundering. In *Proceedings of wavelet analysis and pattern recognition* (ICWAPR '08). 209–215. Hong Kong: IEEE.

Oatley, G. C., B. W. Ewart, and J. Zeleznikow. 2006. Decision support systems for police: Lessons from the application of data mining techniques to 'soft' forensic evidence. *Artifical Intelligence and Law* 14: 35–100.

Priddy, K. L., and P. E. Keller. 2005. Artificial neural networks: An introduction. Washington, DC: SPIE-The International Society for Optical Engineering.

Ribaux, O., and P. Margot, P. 1999. Inference structures for crime analysis and intelligence: The example of burglary using forensic science data. *Forensic Science International* 100:193–210.

Shearer, C. 2000. The CRISP-DM Model: The new blueprint for data mining. *Journal of Data Warehousing* 5:13–22.