Expert Systems with Applications 41 (2014) 2052–2064

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Genetic algorithm-based heuristic for feature selection in credit risk assessment

Stjepan Oreski *, Goran Oreski

Bank of Karlovac, I.G.Kovacica 1, 47000 Karlovac, Croatia

A R T I C L E   I N F O

Keywords:
Artificial intelligence
Genetic algorithms
Classification
Credit risk assessment
Incremental feature selection
Neural network

A B S T R A C T

In this paper, an advanced novel heuristic algorithm is presented, the hybrid genetic algorithm with neural networks (HGA-NN), which is used to identify an optimum feature subset and to increase the classification accuracy and scalability in credit risk assessment. This algorithm is based on the following basic hypothesis: the high-dimensional input feature space can be preliminarily restricted to only the important features. In this preliminary restriction, fast algorithms for feature ranking and earlier experience are used. Additionally, enhancements are made in the creation of the initial population, as well as by introducing an incremental stage in the genetic algorithm. The performances of the proposed HGA-NN classifier are evaluated using a real-world credit dataset that is collected at a Croatian bank, and the findings are further validated on another real-world credit dataset that is selected in a UCI database. The classification accuracy is compared with that presented in the literature. Experimental results that were achieved using the proposed novel HGA-NN classifier are promising for feature selection and classification in retail credit risk assessment and indicate that the HGA-NN classifier is a promising addition to existing data mining techniques.

1. Introduction

Credit risk is one of the most important issues in the banking industry; therefore, credit risk assessment has gained increasing attention in recent years (Akkoç, 2012; Danenas et al., 2011; Finlay, 2011; Tsai, Lin, Cheng, & Lin, 2009). Until a few years ago, the body of research on consumer credit risk assessment was quite sparse. Quantitative consumer credit risk assessment models were developed much later than those for business credit, mainly due to the problem of data availability. Data were limited to the databases of financial institutions. Currently, some data are publicly available in several countries, and financial institutions and researchers have developed many different quantitative credit scoring techniques (Šušteršic, Mramor, & Zupan, 2009). Still, there is no standard set of attributes or indicators that would exist in all credit institutions and on the basis of which the classification of retail customers in terms of credit worthiness could be conducted. Therefore, it is necessary to use all of the available data and information, methods and algorithms for feature selection and the precise classification of clients.

Forced by a crisis, banks are exposed to challenges in finding new ways of doing business that must be less risky and entirely efficient and profitable. They are forced by the crisis because the crisis has revealed the level of risks that are embedded in the banking business. In retail, risks were very often taken with no exact estimate of their degree and possible consequences. The large number of decisions that are involved in the consumer lending business make it necessary to rely on models and algorithms rather than on human discretion and to base such algorithmic decisions on “hard” information (Khandani, Kim, & Lo, 2010). In terms of the high growth of the economy, banks achieved high rates of profit for their owners and were not exposed to the challenges of finding substantially new ways of doing business.

The crisis has significantly reduced the profit margin; many banks have run into difficulties, and some have gone into bankruptcy. Investors have become increasingly cautious and not willing to invest their capital in such troubled banks, and regulators require banks to strengthen their capital (BIS, 2011), raising their resistance to such emergencies. State governments and international organisations have become involved in rescuing the situation and preventing even larger consequences of the crisis.

The crisis was deep because one of the fundamental causes of the crisis was the way that banks functioned. Few exact methods have been used in assessing the retail risk, and taking collateral for borrowed funds has been used as a major surrogate. When it became obvious that this collateral was not as valuable as its prior assessed value and that the credit risk of the clients was not appropriately assessed, the losses became inevitable. We should work on

---

* Corresponding author. Tel.: +385 98 246 328; fax: +385 47 614 306.
E-mail addresses: stjepan.oreski@kaba.hr (S. Oreski), goran.oreski@gmail.com (G. Oreski).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.09.004

all of the causes of the crisis and even on the way that banks do business, primarily in terms of taking a risk when lending funds. Operations should be faster, less risky, more exact and based on data. Banks should use the capital at their disposal in a better way. This capital is not only in terms of money but is also in terms of the data collected in their databases. The capital in the form of customer data should be managed better by the banks. It should be transformed into knowledge and ultimately money.

The data in the databases can be used for credit risk assessments, but these data are commonly high dimensional. Irrelevant features in a training dataset could produce less accurate results in the classification analysis. Feature selection is required to select the most significant features to increase the predictive accuracy, speed and scalability. In difficult learning problems, such as in credit risk assessment, using the appropriate set of features is critical for the success of the learning process and therefore, by itself, is an important issue. Hence, we investigate the possibilities for feature selection methods that provide increased accuracy and scalability of the algorithms and that enable incremental feature selection. A novel and efficient hybrid classifier is designed here.

The present research is focused on the genetic algorithm (GA) and its capabilities for enhancement. The enhancement of the genetic algorithm involves the prevention of spending time in exploring irrelevant regions of the search space. Therefore, the theme of this paper is the advanced heuristic algorithm creation by the hybridisation of the genetic algorithm with some of the filter techniques. The novel classifier, called HGA-NN, is composed of fast filter techniques, a hybrid genetic algorithm (HGA) and an artificial neural network. Research was conducted on solving the problems of feature selection and classification when assessing retail credit risks.

The remaining sections of this paper are organised as follows. Section 2 describes the problem of feature selection to be studied in the paper and reviews the previous literature related to the problem. A brief overview of the genetic algorithm design is given in the third section. Section 4 describes the experimental design and model development. Section 5 discusses the experimental results with performance evaluation and comparison. Section 6 concludes this paper and provides guidelines for future work.

2. Problem statement and literature review

Feature selection is a pre-processing technique that is commonly used on high-dimensional data, and its purposes include reducing the dimensionality, removing irrelevant and redundant features, facilitating data understanding, reducing the amount of data needed for learning, improving the predictive accuracy of the algorithms, and increasing the interpretability of the models (Oreski, Oreski, & Oreski, 2012). Feature selection is the problem of choosing a small subset of features that ideally is necessary and sufficient for describing the target concept (Kira & Rendell, 1992). When the feature selection is poorly performed, it could lead to problems that are associated with incomplete information, noisy or irrelevant features, and not the best set of features, among other problems. The learning algorithm that is used is slowed down unnecessarily due to the large number of dimensions of the feature space, while also experiencing lower classification accuracies due to learning irrelevant information.

The problem of the selection of m features from a set of n features can be solved with different algorithms. From the perspective of the processor time necessary for solving the problem, the computational complexity of this problem is $\binom{n}{m}$ and belongs to the class of NP problems. For larger dimensions, these problems cannot be solved by means of exhaustive search or simple heuristics. In recent years, various feature selection algorithms (techniques) have been proposed. Some of them will be mentioned below.

Aha and Bankert (1996) report positive empirical results with forward and backward sequential feature selection algorithms. They show that feature selection improves the performance of classifiers and provides evidence that wrapper models outperform filter models. Danenas et al. (2011) applied feature selection for datasets by using a correlation-based feature subset selection algorithm with Tabu search for search in attribute subsets. Jin et al. (2012) proposed the attribute importance measure and selection method based on attribute ranking. In the proposed attribute selection method, input output correlation is applied for calculating the attribute importance, and then the attributes are sorted in descending order. The hybrid of Back Propagation Neural Network (BPNN) and Particle Swarm Optimisation (PSO) algorithms is also proposed. PSO is used to optimise weights and thresholds of BPNN for overcoming the inherent shortcoming of BPNN. Their experimental results show that the proposed attribute selection method is an effective preprocessing technique.

Piramuthu (2006) considers decision support tools for credit-risk evaluation from a machine learning perspective. He discusses a few means of improving the performance of these tools through data preprocessing, specifically through feature selection and construction. He stated simply that one must take data and/or problem characteristics as well as the suitability of a given algorithm to obtain better performance. Performance, in this context, depends on at least two different entities: the algorithm and the dataset.

All of these feature selection techniques can be divided into three groups: filter, wrapper and hybrid techniques. The filter techniques rely on the general characteristics of the data to evaluate and select attribute subsets without involving a classification algorithm. One advantage of filter techniques is that because they do not use the classification algorithm, they are usually fast and therefore suitable for use with large datasets. Additionally, they are easily applicable to various classification algorithms. The wrapper techniques first implement an optimising algorithm that adds or removes attributes to produce various subset attributes and then employ a classification algorithm to evaluate this subset of attributes. The wrapper techniques are known to be more accurate compared to the filter techniques, and they are computationally more expensive. Because the classification algorithm is called repeatedly, wrapper techniques are slower than filter techniques and do not scale up well to large, high-dimensional datasets. The hybrid techniques attempt to take advantage of the filter and wrapper techniques by exploiting their complementary strengths (Jin & et al., 2012).

Hybrid techniques are usually a combination of filter and wrapper techniques and are designed to trade the accuracy with the computational speed by applying a wrapper technique to only those subsets that are preselected by the filter technique (Jin et al., 2012). The strategies used for searching the feature space in hybrid techniques are very different. Because of the time complexity of the problem, meta-heuristics are often used. One of the meta-heuristics is GAs. The advantage of GAs compared with other search algorithms is that more strategies can be adopted together to find good individuals to add to the mating pool in a GA framework, in both the initial population phase and the dynamic generation phase (Pezzella, Morganti, & Ciaschetti, 2007). Recently, various variants of GAs have been proposed.

Yang, Li, and Zhu (2011)) describe an improved genetic algorithm for optimal feature subset selection from a multi-character feature set (MCFS). They divide the chromosome into several segments according to the number of feature groups in MCFS for local management. A segmented crossover operator and a segmented mutation operator are employed to operate on these segments to avoid invalid chromosomes. The probability of crossover and mutation are adjusted dynamically according to the generation number and the fitness value. As a result, they obtain strong searching ability at the beginning of the evolution and achieve accelerated convergence along the evolution.

Li et al. (2011) focus on the strategies of generating the initial population of a genetic algorithm and examine the impact of such strategies on the overall GA performance in terms of the solution quality and the computational time. Their initialisation approach first uses a greedy algorithm to quickly generate a high-quality solution with low computational complexity and then uses this solution as a seed to generate a set of solutions as the initial GA population, which is then used in a hybrid GA to test the performance of the proposed approach. To enhance the quality of the initial solution, Zhang et al. (2011) designed a new initial assignment method to generate a high-quality initial population that integrates different strategies to improve the convergence speed and the quality of the final solutions. Maaranen et al. (2004) study the use of quasi-random sequences in the initial population of a genetic algorithm. They conclude that different initial populations of a genetic algorithm have an effect on the final objective function value and the total number of generations used.

However, there is no study that compiles the initial population of a genetic algorithm based on the domain experts' knowledge and the results of the filter techniques for the feature subset selection, which is performed with the following: Information gain (IG), Gain ratio, Gini index and Correlation, and the remaining initial population is generated randomly. According to this arrangement, the hypothesis H1 is defined as follows:

H1: The incorporation of the preliminary feature selection and the incremental stage to a GA-based algorithm combined with the effects of the GA initial population generating strategy results in a statistically significant enhancement of the average classifier accuracy with a 99% confidence level.

It is important to note that the improvement even of a fraction of a percent is large enough to be scientifically and practically interesting. To test our hypothesis and to assess the accuracy of our technique, experiments were conducted on two real-world credit datasets, one collected at a Croatian bank and another selected from the UCI database.

3. Genetic algorithm design

Genetic algorithms (GAs) can be described as a heuristic search and optimisation technique that is inspired by natural evolution (McCall, 2005), which transposes the notions of natural evolution to the world of computers and imitates natural evolution. GAs were initially introduced by John Holland for explaining the adaptive processes of natural systems and for creating new artificial systems that work on similar bases. In Nature, new organisms adapt to their environment through evolution. Genetic algorithms evolve solutions to the given problem in a similar way (Renner & Ekárt, 2003).

The implementations of genetic algorithms can significantly differ in the way of constructing a new population. Some implementations create a separate population of new individuals in every generation by applying genetic operators (shown in Fig. 1). Other implementations extend the current population by adding new individuals and then create the new population by omitting the least fit individuals (shown in Fig. 2). There are GAs that do not use generations at all and instead have continuous replacement. According to the method of creating a new population, the GA adapts other operators, especially the selection operator, as well the position of the fitness calculation of the individuals.

On the one hand, a simple core of the genetic algorithm is available; on the other hand, we must consider the specific problem that we want to solve. Presentation, control mechanisms, fitness function, the method of initialisation and genetic operators should also be appropriately adapted to the problem. A GA has a specific strength because standard components can be re-used, with adaptation to many different situations, which eases the implementation. The implementation of the genetic algorithm used in this research basically follows the algorithm described in Fig. 2. A detailed description of the genetic algorithms and guidance on when it is most appropriate to use them can be found in Michalewicz (1998) and Mitchell (1996).

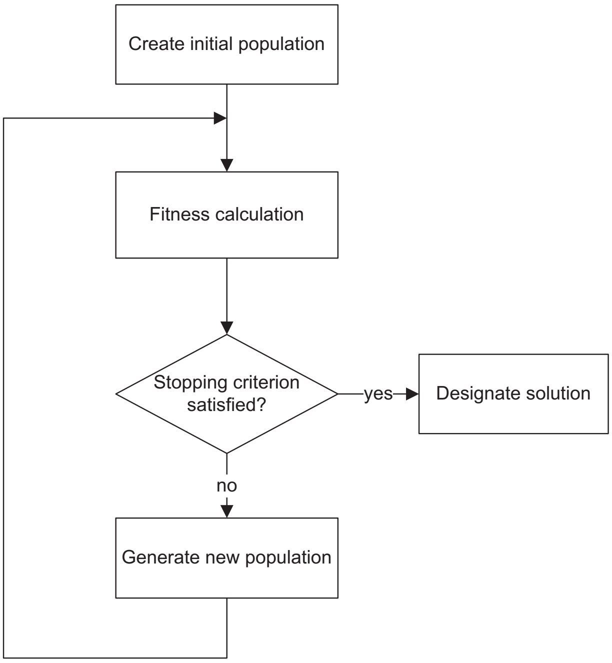

Fig. 1. Algorithmic description of the genetic algorithm with the creation of a separate population.

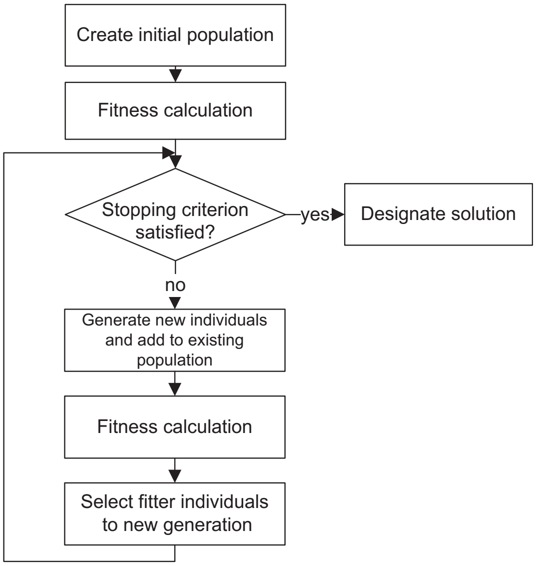

Fig. 2. Algorithmic description of the genetic algorithm with the extension of the current population.

4. Model development

A GA as the evolutionary optimisation technique starts with an initial population (initial solutions), moves toward a global optimal solution and stops searching when the stop conditions are satisfied. The ordinary approach begins with a random initial population and then evolves from one generation to another as the individuals undergo crossover and mutation (Figs. 1 and 2). Our approach is based on the following basic hypothesis: when a priori information about the potentially attractive areas is available, then the initial population of the GA can be generated in such a way that the attractive areas of the feasible region must be covered with a set of points and the dimensionality of the problem can be reduced to those features that form attractive areas. In this way, the attractive area (the region of attraction) of a global minimum is defined as the largest set of points, such that for any starting point from that set, the infinitely small step of the gradient decent algorithm will converge to the global minimum (Maaranen et al., 2004). The results of earlier experience and the results of the fast filter techniques are considered to be a priori information about the attractive areas. In fact, this a priori information comprises initial solutions that are included in the initial population of the GA, and the remainder of the initial population is filled randomly. The proposed algorithm can be seen as a unified whole, as is shown in Fig. 3, or it can be viewed by execution phases, as is shown in Fig. 4.

4.1. The search space reduction

The aim of the first stage is preliminary to restricting the search space. According to Figs. 3 and 4, the concept is to prevent the genetic algorithm from consuming time in exploring irrelevant regions of the search space. The search space reduction is accomplished in two phases: (1) the generation of initial solutions, and (2) the generation of a reduced feature subset.

The generation of the initial solutions is performed by means of the filter techniques for feature ranking: Information gain, Gain ratio, Gini index and Correlation. In addition, solutions offered by domain experts on the basis of their earlier experience are incorporated into the initial solutions. If this type of knowledge does not exist, then we can incorporate the previously known solution, if it exists. In the second stage, the initial solutions will be embedded into the initial population of the HGA.

The generation of the reduced feature subset represents the union of all of the features that appear in the previously known solutions and in the results obtained from the fast filter techniques, i.e., this set represents the union of all of the features that appear in the initial solutions. Only the features that appear in the results of the filter techniques and in the previously known solutions will appear at the entrance to the HGA and act as the reduced feature set. The reduced feature set is further refined in the second and the third stage of the HGA-NN. In this way, these restrictions to the search space can make the feasible space very small compared with the entire search space.

The preliminary restriction of the features has a very important role in the HGA-NN design. Using a large number of different techniques often results in selecting a large number of features.

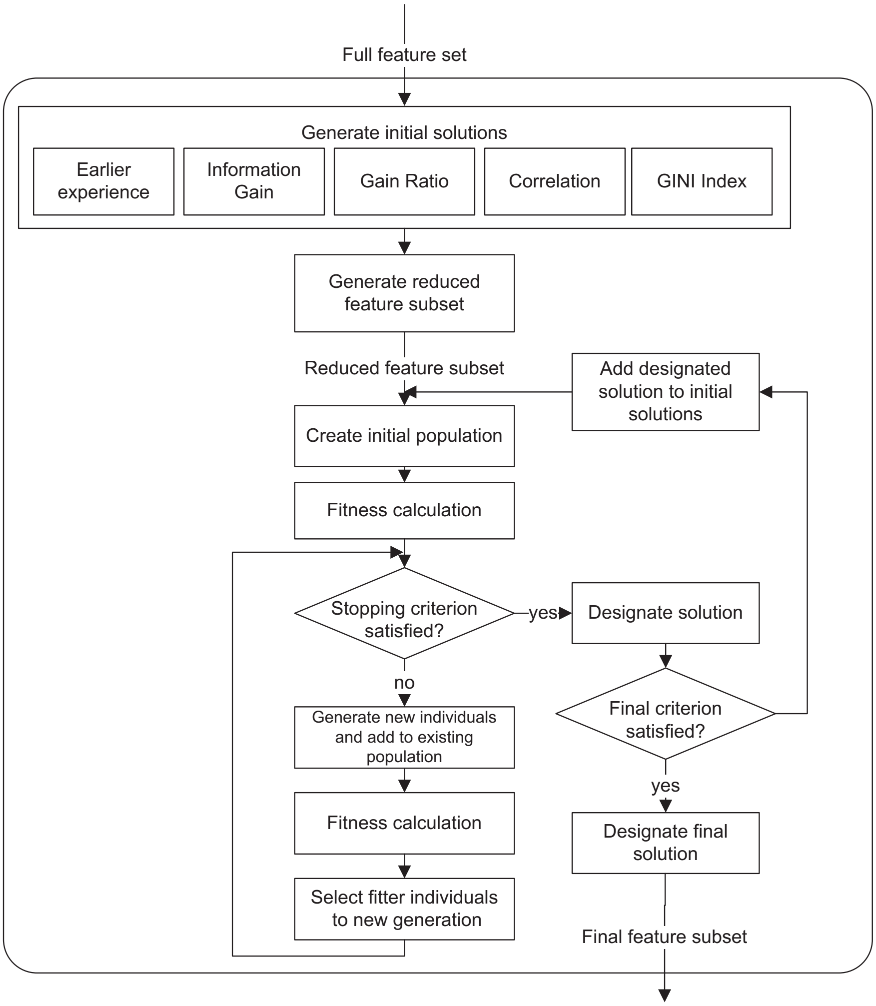

Fig. 3. Flowchart of the hybrid genetic algorithm.

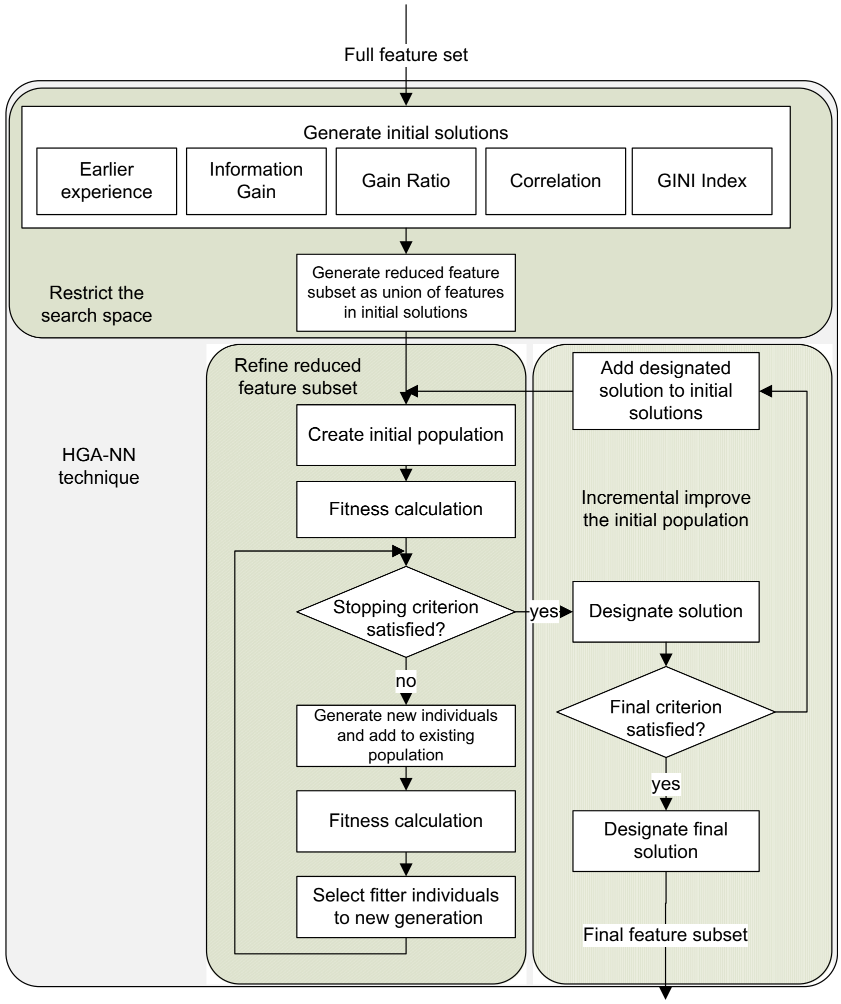

Fig. 4. The hybrid genetic algorithm divided into stages.

Through having a low selection pressure, the basic task of this algorithm phase is not fulfilled. Then, the scalability of the algorithm on feature selection problems with greater cardinality is questionable.

In contrast, using few and inadequate fast filter techniques could lead to the omission of important attributes. Given that data characteristics essentially affect feature selection technique performances, a good knowledge of the data characteristics is an essential prerequisite for a good selection of fast techniques in the preliminary stage.

4.2. Refining of the reduced feature subset

It has been shown that a simple combination of the best individual features selected by the filter selection techniques does not necessarily lead to a good classification performance (Kohavi & John, 1997). In other words, the redundancy among the features could cause the degradation of the classification performance. As a result, the reduced feature subset has been additionally refined.

This refinement is made by the HGA. The standard GA has been improved so that it can accept, as part of its own initial population, the initial solutions generated by the other techniques and the domain's experts. This improvement has enabled also the introduction of the third stage of the algorithm.

The algorithm shown in Fig. 3 was made using the Rapid Miner 5.1.15 tool with the parameters shown in Table 1. Some improvements of this tool were necessary, as mentioned earlier, to execute our algorithm. In Table 1, the parameters for the HGA and some parameters for the NN, the component of the GA-NN and the HGA-NN, can be observed. Parameters for the NN are not changed throughout our experiment; in this way, the NN will not influence the resulting differences among the experimental algorithms. The parameters that are changed throughout conducting the experiments are marked with a ✓ in the second column of the table, and all of them are related to the GA component of the experimental algorithms.

4.3. Incremental stage

As Goldberg and Holland (1988) state, genetics-based classifiers rely on competition to resolve conflicts. This property makes it possible for the system to operate incrementally, testing new structures and hypotheses while steadily improving its performance. This approach is especially important if the process converges quickly to the local extreme (local optimum). Because further possibilities for the standard genetic algorithm to improve solutions are very small, a new reconstruction of the genetic material is necessary. This role has incremental stages. The incremental stages allow the reconstruction of the genetic material and contribute to the diversity of the genetic material.

The influence of the initial population on the convergence rate to a local optimum can be observed as follows: If we initially select relatively well-adjusted individuals who belong to different

Table 1
Summary of the HGA-NN parameters.

Parameter	Changed	Setting
Population initialisation		
Population size		50
Initial probability for a feature to be switched on (p_initialise)	✓	0.5
Maximum number of features		12
Minimum number of features		5
Reproduction		
Fitness measure		Accuracy
Fitness function		Neural network
The type of neural network		Multilayer feed-forward network
Network algorithm		Backpropagation
Activation function		Sigmoid
The number of hidden layers		1
The size of the hidden layer		(number of features + number of classes)/2 + 1
Training cycles		[300; 600]
Learning rate		[0.3; 1.0]
Momentum		[0.2; 0.7]
Selection scheme	✓	Roulette wheel
Tournament size	✓	–
Dynamic selection pressure		Yes
Keep best individual		Yes
Mutation probability		1/number of features
Crossover probability		0.9
Crossover type	✓	Uniform
The condition for the completion		
Maximal fitness		Infinity
Maximum number of generations		50
Use early stopping		No

attractive areas, the crossing does not converge too fast to the local optimum. In contrast, in conditions when there are initially relatively good individuals who belong to the same attractive zone, the process converges quickly to a local optimum. The genetic material of a super-individual would quickly dominate the population, and premature convergence would ensue (Michalewicz, 1998). In this way, a certain amount of the genetic material that could be useful and belongs to different areas is lost because it is contained in the bad individuals (Bäck, Fogel, & Michalewicz, 1997).

For this reason, we do not know in advance what type of data are present, and it is necessary to have the capability of reconstructing the genetic material and altering the algorithm's parameters to improve the model performance. It is important to note that, after each stage of the hybrid algorithm, we can dynamically change the conditions for the next stage of the algorithm. Thus, before every new execution of stage 2 (refine the reduced feature subset), we can change the initial population, the selection scheme, the crossover type, the mutation rate or any other parameter of the HGA and control the results of these changes.

5. Empirical analysis

The performances of the proposed HGA-NN classifier are evaluated using the real-world credit dataset collected at a Croatian bank, and the findings are further validated on another real-world credit dataset selected from the UCI database. In this work, the empirical analysis is divided into: (1) Experiment 1 with the Croatian dataset and (2) Experiment 2 with the German dataset.

5.1. Experiment 1: Croatian dataset

This section is divided into four subsections. In Section 5.1.1, the nature and characteristics of the dataset are described. Section 5.1.2 discusses the parameter setup that is required for the experiment and the achieved results. The comparative performances of the algorithms are demonstrated in Section 5.1.3, with a discussion. The statistical analysis of the achieved results is presented in Section 5.1.4.

5.1.1. Description of the dataset

The Croatian credit dataset was collected at a Croatian bank and covered the period from September 2004 to September 2011. The total number of randomly selected instances in the dataset is 1000, including the 750 who successfully fulfilled their credit obligations, i.e., good credit customers, and 250 who were late in performing their obligations and therefore are placed in a group of bad credit customers. The total number of features is 35, including 33 regular features and 2 (id, label) special features. The regular features contain 21 integer and 12 real values. The class label feature is binominal. There are no missing feature values in the dataset.

The features were divided into five main groups: (i) basic characteristics; (ii) payment history (monthly averages); (iii) financial conditions; (iv) delinquency history; and (v) past credit experiences. These features refer to a client's gender and age, the credit purpose, the credit amount, the number of existing credits at the same bank, the credit history with the bank before the loan was granted, the instalment rate in the percentage of the disposable income and detailed data on the accounts balances and transactions with the bank. The full list of features with their explanations and descriptive statistics and a description of the data collection process for the data sample can be found in Oreski et al. (2012).

5.1.2. Experimental results

As was described in the Problem statement and in the literature review section, the study findings are organised around the following two main issues: (1) the impact of the search space reduction and (2) the incremental stage of the overall classifier performance. The combined effects of the strategy of generating the initial population of a GA to the point of obtaining the results are also addressed within the context of the two main issues.

Based on a number of experiments, it was found that the estimated accuracy began to fall after reducing the number of features below 12. For this reason, the top 12 features, shown in Table 2, are selected by using the aforementioned techniques. Two special features, the first (id) and the last (label), are omitted from the table. Once the features for every technique have been selected, the features that appear in any selection are chosen in the reduced feature subset.

The features that appear in more than half of the other techniques are relevant for the Voting technique. The specificity of the Voting technique, in this context, is that this technique gives only one initial solution and has no impact on the reduced feature subset. From Table 2, it can be observed that we have reduced the original set of features from 33 to 21 features in the phase of the search space reduction. In this way, as was mentioned earlier, a reduced version of a given problem is obtained with the assumption that: (1) the important features are retained, and (2) the resulting reduced search space can improve the scalability of the technique and the quality of the final solution.

With the aim of comparing the quality and the precision of our new HGA-NN technique, parallel tests are performed with the existing GA-NN technique on the same dataset. The GA-NN technique does not perform the search space reduction and does not have an incremental phase. In fact, the second phase in the HGA-NN technique called "Refining the reduced feature subset" is derived from the GA-NN, the technique in the comparison. In this way, the comparison will show if the added phases, the first and the third phase, in the HGA-NN technique bring improvements. We again note that the initial parameters concerning the common

Table 2
The initial solutions and the reduced feature subset for the Croatian dataset.

Original feature set	Feature selection technique					Reduced feature subset
	Current solution	GINI and IG	Gain ratio	Correlation	Voting	
Group1						
att2		✓		✓		✓
att3						
att4						
att5	✓					✓
att6				✓		✓
att7	✓	✓		✓	✓	✓
att8			✓	✓		✓
att9						
att10	✓		✓	✓	✓	✓
att11						
att12		✓				✓
Group2						
att13		✓	✓			✓
att14						
att15		✓	✓			✓
att16	✓					✓
att17	✓	✓	✓		✓	✓
att18	✓		✓			✓
att19	✓					✓
att20						
att21						
att22						
att23	✓	✓	✓	✓	✓	✓
att24	✓	✓	✓		✓	✓
Group3						
att25						
att26						
att27		✓	✓	✓	✓	✓
att28	✓	✓	✓	✓	✓	✓
Group4						
att29				✓		✓
att30		✓	✓	✓	✓	✓
att31	✓	✓	✓	✓	✓	✓
Group5						
att32						
att33						
att34	✓			✓		✓

parts of the techniques, the genetic algorithm and the neural networks, were set to the same values in both techniques. The reason for this choice is primarily from a simple comparison of the results.

The results of our experiments are presented in Tables 4 and 5. Every value in Tables 4 and 5 present an average prediction accuracy of the mentioned techniques obtained on a set of 1000 individuals. The accuracy is calculated by using the 10-fold cross-validation technique because a simple validation technique often estimates the true error rate too optimistically (Malhotra & Malhotra, 2003). In our case, the 10-fold cross-validation technique uses stratified sampling because the percentage of bad loans is small compared to the performing loans. Stratified sampling builds random subsets and ensures that the class distribution in the subsets is (almost) the same as in the whole example set.

After the initial setting of the parameters shown in Table 1, for the vertical comparison of the results to make sense, most of the parameters did not change from one execution to another. The vertical comparison of the results is the comparison within the same technique that occurred due to changes in the parameters. The changed parameters are highlighted in Table 3. They were also being placed on the same values in both techniques for the same execution.

As can be seen from Table 3, there were 11 executions. Every execution corresponds to one scenario of testing. In each scenario, we made some changes to test many different options. The meanings of the parameters as well as their types of values can be found in Bäck et al. (1997) and Michalewicz (1998) with the exception of the unique selection. The unique selection is our own selection scheme. With the unique selection, the individuals are ordered according to their fitness value, and the ones that have the best fit but are still distinct are selected. The number of the selected individuals is less than or equal to the population size. If there are not enough distinct individuals, then the population is supplemented with the best fit individuals.

The parameter called “Tournament size” has values in the table only when the selection scheme is tournament; in the other selection schemes, the value of this parameter has no meaning. Additionally, the scenarios from 1 to 8 did not include the incremental feature of the HGA; this feature is included in scenarios from 9 to 11. For every scenario (execution), the results are shown in Tables 4 and 5 for 9 different generations of a genetic algorithm. In this way, we obtain 99 average prediction values, which are expressed in terms of the percentage accuracy for each technique.

Comparing the changed parameters shown in Table 3 with the results of execution shown in Tables 4 and 5, we can see that the change in the parameter called p_initialise, which means the initial probability of a feature to be switched onto the initial solution, has had a direct impact on the result of the initial (first) generation of the third and subsequent execution of the GA-NN technique. At the same time, this change did not have an impact on the results of the HGA-NN technique. It is obvious that the initial solutions embedded in the initial population of the HGA-NN are better than any

Table 3
The parameter changes for the different executions.

Execution	Parameter			
	p_initialise	Selection scheme	Tournament size	Crossover type
1	0.5	Roulette wheel	–	Uniform
2	0.5	Tournament	0.1	Uniform
3	0.6	Tournament	0.05	Uniform
4	0.6	Stochastic	–	One point
5	0.6	Boltzmann	–	Uniform
6	0.6	Cut	–	Uniform
7	0.6	Unique	–	Uniform
8	0.6	Unique	–	One point
9	0.6	Tournament	0.05	Uniform
10	0.6	Tournament	0.05	One point
11	0.6	Cut	–	One point

other solution in the initial population that was generated randomly.

The first change in the first generation of the HGA-NN occurred in the 9-th execution. This change occurred because the result of the 8-th execution of the HGA-NN was added to the initial solutions for the 9-th execution. In other words, in the 9-th execution of the HGA-NN, a feature called the incremental stage was activated. Along with an impact on the result of the first generation, the incremental stage has had an impact on the final result of the HGA-NN technique. This point was the first time that the accuracy of the HGA-NN reached the value of 83.4%. Although we added the result of the preceding execution in the initial population of the 10-th and 11-th execution of the HGA-NN, the final results are not improved. After 2 generations without any improvement in the performance, the process was stopped. Comparing the results before and after the activation of the incremental stage, it can be seen in Table 5 that there is a beneficial effect of this activation on the classification accuracy, although the HGA-NN technique already achieved on average better results than the GA-NN.

5.1.3. Comparison and discussion of the results

To estimate the efficiency of the HGA-NN algorithm and to quantify the impact of the search space reduction and the incremental stage to the overall classifier performance, we compare its accuracy with the accuracy of the GA-NN algorithm. A simple comparison of these two feature selection techniques is possible because the same initial parameters and the same classification and validation methods are used. Summary results of exhaustive experiments are shown in Table 6. From these results, it is observed that, in all 9 measured generations, from 1 to 50, the HGA-NN achieves better results than the GA-NN.

The trends are even more obvious from Figs. 5 and 6, which show the same results observed from two different positions, with two different aspects. Therefore, the scales on the X-axis are different. The measuring scale is set to be balanced with respect to the anticipated results of the measurement. The balance, which is a nearly straight line, is achieved, as can be seen from Fig. 5. Additionally, Table 6 and Fig. 5 show that the difference between the

Table 4
The prediction accuracy of the GA-NN technique expressed in % for the Croatian dataset.

Execution	The generations of the GA-NN								
	1	2	3	5	8	10	20	30	50
1	79.7	80.5	80.5	80.8	81.6	81.8	81.8	81.8	81.8
2	79.7	80.4	81.1	81.6	82.1	82.1	82.4	82.4	82.4
3	80.5	81.3	81.3	81.3	81.4	81.8	81.9	82.5	82.5
4	80.5	80.5	80.9	81.7	81.8	82.2	82.2	82.2	82.2
5	80.5	80.6	80.9	80.9	81.1	81.1	81.9	81.9	81.9
6	80.5	80.6	81.3	82.2	82.2	82.2	82.9	83.1	83.1
7	80.5	80.6	81.3	81.5	82.2	82.2	82.2	82.3	82.7
8	80.5	81.0	81.0	81.2	82.1	82.1	82.3	82.5	82.6
9	80.5	81.3	81.3	81.3	81.4	81.8	81.9	82.5	82.5
10	80.5	80.9	81.5	81.5	81.9	82.6	82.8	82.8	82.8
11	80.5	81.1	81.3	81.5	81.5	81.5	82.2	82.2	82.5
Mean	80.35	80.8	81.13	81.41	81.75	81.95	82.23	82.38	82.45
Std. dev.	0.324	0.332	0.283	0.383	0.378	0.406	0.364	0.371	0.378

Table 5
The prediction accuracy expressed in % for the Croatian dataset using the HGA-NN technique.

Execution	The generations of the HGA-NN								
	1	2	3	5	8	10	20	30	50
1	81.3	81.3	81.3	81.3	82.2	82.3	82.3	82.3	82.3
2	81.3	81.4	82.0	82.6	82.8	82.8	82.8	82.8	82.8
3	81.3	81.3	82.4	82.4	82.4	82.4	82.7	83.0	83.1
4	81.3	82.0	82.0	82.2	82.2	82.2	82.4	82.5	82.5
5	81.3	81.3	81.3	81.3	81.5	81.5	81.7	81.8	81.8
6	81.3	81.3	82.5	82.5	82.5	82.5	82.5	82.5	82.9
7	81.3	81.3	81.3	81.7	81.7	81.7	82.4	82.4	83.0
8	81.3	81.3	81.4	81.8	81.8	82.5	82.5	82.7	83.1
9	81.7	82.0	82.2	82.3	82.9	82.9	82.9	82.9	83.4
10	82.0	82.5	82.5	82.8	82.9	82.9	83.4	83.4	83.4
11	82.0	82.1	82.1	82.3	82.4	82.5	82.9	83.0	83.4
Mean	81.46	81.62	81.91	82.11	82.30	82.38	82.59	82.66	82.88
Std. dev.	0.291	0.442	0.495	0.511	0.48	0.451	0.432	0.43	0.508

Table 6
The comparison of the average prediction accuracy of both techniques for the Croatian dataset.

Technique	The generations of the GA								
	1	2	3	5	8	10	20	30	50
GA-NN	80.35	80.80	81.13	81.41	81.75	81.95	82.23	82.38	82.45
HGA-NN	81.46	81.62	81.91	82.11	82.30	82.38	82.59	82.66	82.88

average prediction accuracy of the two techniques in the first 10 generations of the GAs narrows, while the difference in the prediction accuracy remains the same between the 10-th and 50-th generations of the GAs. We attribute this trend to the fact that the HGA-NN technique received several good initial solutions from which it did not find a much better combination than the GA-NN did. At the same time, in the GA-NN technique, the genetic algorithm started with a randomly generated initial population and found a combination of genes that improved the precision of the objective function better than the HGA-NN. Despite the initially slower improvement in the objective function value, the HGA-NN algorithm throughout all of the generations achieves better results.

In terms of a percentage, it appears that the average prediction accuracy of the GA-NN technique in the first 10 generations improved by 1.6 percentage points, while at the same time, the average prediction accuracy of the HGA-NN technique improved by 0.92 percentage points. The average prediction accuracy of both techniques improved between the 10-th and the 50-th generations by 0.5 percentage points.

From Fig. 6, it is clear that the values of the objective functions after the 10-th generation are growing very slowly, which is consistent with the earlier remark that the values of the objective functions over the last 40 generations, from the 10-th to 50-th generations, grew less than in only the first 4 generations.

From the results presented in Table 6, Figs. 5 and 6, we can infer that, in solving the feature selection problem for the credit dataset collected at a Croatian bank, the performance of the HGA-NN technique is better than the GA-NN technique regarding the average accuracy of all of the generations, including the accuracy of the final solution. This inference is supported by statistical tests.

5.1.4. Statistical analysis

We have tested the proposed novel HGA-NN technique with the GA-NN technique using the paired *t*-Test for their summarised average accuracy scores shown in Table 6. The null hypothesis is that there is no difference between the average accuracy of these two techniques. In comparing the paired *t*-Test with a *t*-value of 6.8136 and 8 degrees of freedom, we rejected the null hypothesis because the calculated *t*-value is greater than the critical two-tail tabulated value 3.3554 with the chosen level of significance 0.01. The paired *t*-Test assumes that we have sampled our pairs of values from a population of pairs in which the difference between pairs follows a Gaussian distribution (Myers & Well, 2003). For confirmation, we tested this assumption with a normality test. Because the D'Agostino and Pearson omnibus normality test p-value = 0.6137 is greater than $a = 0.05$ and the Shapiro–Wilk normality test p-value = 0.59 is greater than $a = 0.05$, there is not sufficient evidence to support the claim that we have sampled our pairs of values from a population of pairs that is different from a Gaussian distribution (see Table 7).

Together with the paired *t*-Test, we have tested the proposed novel HGA-NN technique with the GA-NN technique using the Wilcoxon matched-pairs signed rank test. This test provides a powerful alternative to the paired *t*-Test when the population of difference scores is symmetrically distributed. This test is only slightly less powerful than the *t*-Test when the data are normally distributed and can be considerably more powerful when the difference scores are symmetrically (but not necessarily normally) distributed with heavy tails (Myers & Well, 2003).

The Wilcoxon matched-pairs signed rank test tests the null hypothesis that the median difference equals 0.0 versus the alternative hypothesis that the median difference is not equal to 0.0, i.e., the null hypothesis is that both algorithms perform equally well. According to the table of exact critical values for the Wilcoxon's matched-pairs signed rank test (Myers & Well, 2003), for the significance level of $\alpha = 0.01$ and $N = 9$, the difference between the classifiers is significant if the smaller of the sums is less than 2. The sum of the ranks for positive differences is $R_{\text{pos}} = 45$, and the sum of the ranks for negative differences is $R_{\text{neg}} = 0$. Because the smaller sum of the ranks is $R_{\text{neg}} = 0$, we can reject the null hypothesis at the 99.0% confidence level and reject the idea that the difference is due to chance, and we conclude instead that the populations have different medians in favour of the proposed novel HGA-NN technique. This conclusion is in accordance with the *t*-Test and the hypothesis H1 of the present research.

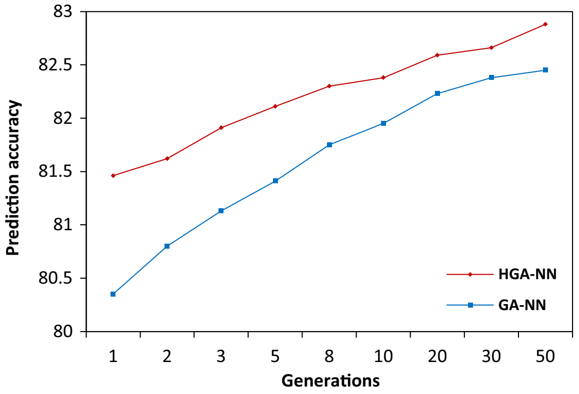

Fig. 5. Comparison of the Croatian dataset results with an interval on the abscissa according to a measurement scale.

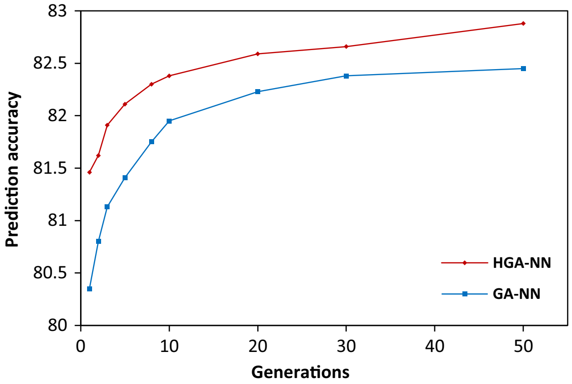

Fig. 6. Comparison of the Croatian dataset results with equal spacing on the abscissa.

Table 7
Statistics tests of the differences between the HGA-NN and the GA-NN for the Croatian dataset.

Test	Level of significance (alpha)	*p*-Value
Paired *t*-Test (two-tailed)	0.01	0.000
D'Agostino and Pearson omnibus normality test	0.05	0.614
Shapiro–Wilks normality test	0.05	0.590
Wilcoxon matched-pairs signed rank test (two-tailed)	0.01	0.004

5.2. Experiment 2: German dataset

For the presented results of the HGA-NN algorithm to be further validated, Experiment 2 was conducted on the publicly available German data set. The results were additionally compared with the other results published in the literature.

5.2.1. Description of the dataset

The German credit dataset is available from the UCI Repository of Machine Learning Databases (Bache & Lichman, 2013), and it is composed of 700 instances of creditworthy applicants and 300 instances of bad credit applicants. The original dataset, in the form provided by Prof. Hofmann, contains, for each applicant, 20 input features that describe the credit history, account balances, loan purposes, loan amounts, employment status, personal information, age, housing and job title. This original dataset had 13 categorical features, some of which have been transformed into a series of binary features so that they can be appropriately handled by the NN. Several ordered categorical features have been left as is and are treated as numerical. The transformed German credit dataset contains 30 regular features of the integer data type and 2 (id, label) special features and can be reached at <http://ocw.mit.edu/courses/sloan-school-of-management/15-062-data-mining-spring-2003/download-course-materials>. All of the features with descriptive statistics are shown in Table 8.

5.2.2. Experimental results

The initial solutions are generated according to the procedure defined in the model development section that has the same feature selection techniques as for the Croatian dataset. After the generation of the initial solutions, the reduced feature subset is generated as the union of the features in the initial solutions. From Table 9, it can be seen that we have reduced the original set of features from 30 to 20 features in the phase of the search space reduction.

According to a number of experiments, the initial parameters of the genetic algorithm were set to the same values as in Experiment 1, except that maximum_number_of_features was set to 16. Additionally, the initial parameters of the neural network were set to the same values as in Experiment 1, except that number_of_training_cycles was set to the interval [30; 100]. All of the initial parameters, from one execution to another, did not change except for the parameters that were highlighted in Table 3.

According to these, for the German credit dataset we also made 99 measures, which are expressed in the percentage of accuracy for each technique. The HGA-NN technique in the 8-th execution reached the best average prediction accuracy of 79.4%, with the described unique selection scheme and the one_point crossover type. The GA-NN technique reached the best result, which was 79.0%, in the 7-th execution, with the unique selection scheme and the uniform crossover type. The best solutions for the HGA-NN and the GA-NN technique were achieved with 12 and 16 features selected, respectively.

Table 8
Transformed German credit dataset with descriptive statistics.

Attr.	Code	Description	Statistics	Range
att1	ID	Observation no.	avg = 500.500 ± 288.819	[1; 1000]
att2	CHK_ACCT	Checking account status	avg = 1.577 ± 1.258	[0; 3]
att3	DURATION	Duration of credit in months	avg = 20.903 ± 12.059	[4; 72]
att4	HISTORY	Credit history	avg = 2.545 ± 1.083	[0; 4]
att5	NEW_CAR	Purpose of credit	avg = 0.234 ± 0.424	[0; 1]
att6	USED_CAR	Purpose of credit	avg = 0.103 ± 0.304	[0; 1]
att7	FURNITURE	Purpose of credit	avg = 0.181 ± 0.385	[0; 1]
att8	RADIO/TV	Purpose of credit	avg = 0.280 ± 0.449	[0; 1]
att9	EDUCATION	Purpose of credit	avg = 0.050 ± 0.218	[0; 1]
att10	RETRAINING	Purpose of credit	avg = 0.097 ± 0.296	[0; 1]
att11	AMOUNT	Credit amount	avg = 3271.258 ± 2822.737	[250; 18424]
att12	SAV_ACCT	Average balance in savings account	avg = 1.105 ± 1.580	[0; 4]
att13	EMPLOYMENT	Present employment since	avg = 2.384 ± 1.208	[0; 4]
att14	INSTALL_RATE	Instalment rate as % of disposable income	avg = 2.973 ± 1.119	[1; 4]
att15	MALE_DIV	Applicant is male and divorced	avg = 0.050 ± 0.218	[0; 1]
att16	MALE_SINGLE	Applicant is male and single	avg = 0.548 ± 0.498	[0; 1]
att17	MALE_MAR_WID	Applicant is male and married or a widower	avg = 0.092 ± 0.289	[0; 1]
att18	CO-APPLICANT	Application has a co-applicant	avg = 0.041 ± 0.198	[0; 1]
att19	GUARANTOR	Applicant has a guarantor	avg = 0.052 ± 0.222	[0; 1]
att20	PRESENT_RESIDENT	Present resident since – years	avg = 2.845 ± 1.104	[1; 4]
att21	REAL_ESTATE	Applicant owns real estate	avg = 0.282 ± 0.450	[0; 1]
att22	PROP_UNKN	Applicant owns no property (or unknown)	avg = 0.154 ± 0.361	[0; 1]
att23	AGE	Age in years	avg = 35.546 ± 11.375	[19; 75]
att24	OTHER_INSTALL	Applicant has other installment plan credit	avg = 0.186 ± 0.389	[0; 1]
att25	RENT	Applicant rents	avg = 0.179 ± 0.384	[0; 1]
att26	OWN_RES	Applicant owns residence	avg = 0.713 ± 0.453	[0; 1]
att27	NUM_CREDITS	Number of existing credits at this bank	avg = 1.407 ± 0.578	[1; 4]
att28	JOB	Nature of job	avg = 1.904 ± 0.654	[0; 3]
att29	NUM_DEPENDENTS	Number of people for whom liable to provide maintenance	avg = 1.155 ± 0.362	[1; 2]
att30	TELEPHONE	Applicant has phone in his or her name	avg = 0.404 ± 0.491	[0; 1]
att31	FOREIGN	Foreign worker	avg = 0.037 ± 0.189	[0; 1]
att32	LABEL	Credit rating	mode = 1 (700), least = 0 (300)	1 (700), 0 (300)

Table 9
The initial solutions and the reduced feature subset for the German credit dataset.

Original feature Set	Feature selection technique					Reduced feature subset
	Current solution	IG	Gain ratio	Correlation	Voting	
att2	✓	✓	✓	✓	✓	✓
att3	✓	✓	✓	✓	✓	✓
att4	✓	✓	✓	✓	✓	✓
att5	✓					✓
att6			✓	✓		✓
att7						
att8		✓		✓		✓
att9						
att10						
att11		✓	✓	✓	✓	✓
att12	✓	✓	✓	✓	✓	✓
att13		✓		✓		✓
att14	✓					✓
att15	✓					✓
att16	✓					✓
att17						
att18						
att19	✓					✓
att20						
att21	✓	✓	✓	✓	✓	✓
att22	✓	✓	✓	✓	✓	✓
att23		✓	✓			✓
att24		✓	✓	✓	✓	✓
att25						
att26		✓	✓	✓	✓	✓
att27						
att28						
att29	✓					✓
att30						
att31	✓		✓			✓

Table 10
The comparison of the average prediction accuracy of both techniques for the German credit dataset.

Technique	The generations of the GA									p-Value[a]
	1	2	3	5	8	10	20	30	50	
GA-NN	75.65	76.88	77.36	77.81	78.00	78.03	78.18	78.37	78.50	
HGA-NN	77.75	78.16	78.25	78.37	78.52	78.55	78.74	78.83	78.90	0.004

[a] The p-Value is for a Wilcoxon matched-pairs signed rank test (alpha = 0.01).

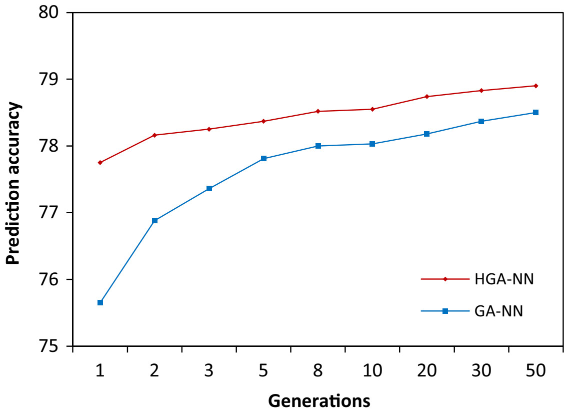

Fig. 7. Comparison of the German credit dataset results with an interval on the abscissa according to a measurement scale.

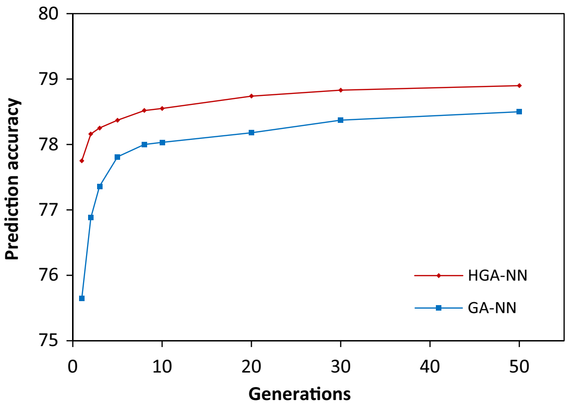

Fig. 8. Comparison of the German credit dataset results with equal spacing on the abscissa.

To additionally validate the findings of Experiment 1, the summary comparative results of Experiment 2 are presented in Table 10 and in Figs. 7 and 8.

Table 10 and Fig. 7 show that the difference between the average prediction accuracy of the two techniques in the first 8 generations of the GAs narrows, while the difference in the prediction

Table 11
The comparison of the results with other techniques (German credit dataset).

Technique	Selected features		Accuracy	
	Average	Std.	Average (%)	Std. (%)
HGA-NN	12.36	1.57	78.90[a]	0.297
GA-NN	13.91	1.81	78.50	0.402
SVM+GA (Huang, Chen, & Wang, 2007)	13.30	1.41	77.92	3.970
GP (Huang et al., 2007)	13.20	2.10	78.10	4.120
NN (Khashman, 2010)	All	–	74.67	–
SVM (Wang et al., 2009)	All	–	76.59	0.328

[a] The best average accuracy.

accuracy remains almost the same between the 8-th and 50-th generations of the GAs.

From Fig. 8, it is clear that the value of the objective function after the 8-th generation is entering the phase of saturation, which is consistent with the findings in Experiment 1. The value of the objective function in the last 40 generations, from the 10-th to 50-th generation, grew less than in only the first 4 generations.

The presented results in Tables 6 and 10 show a clear relation between the HGA-NN and the GA-NN technique and their performance. According to this finding, we can conclude that the impact of the search space reduction and the incremental stage, combined with the effects of the strategy of generating the initial population of a GA to the overall classifier performance, exists and is a positive relationship. This result was consistent for both the Croatian and German data and is supported by statistical tests.

Table 11 compares the results of our proposed HGA-NN technique with other existing techniques, including the GA-NN technique, on the German credit dataset. The average accuracy rates shown here refer to the results of the validation data set. The results obtained using the training data set are not shown. The comparative overview indicates that the proposed GA-based algorithm is an acceptable alternative to optimising both the feature subset and the neural network parameters for the credit risk assessment, and through it, to optimising the set of network weights that provide the best classification performance.

6. Conclusions and future work

With a growing competition and risks, as well as a significantly reduced profit margin in the retail credit industry, operations should be faster, less risky, more exact and based on data. To meet these requirements, the credit institutions need more sophisticated methods for assessing credit risks. Feature selection is one of the challenging issues in the creation of this sophisticated method. In this paper, a new advanced hybrid feature selection heuristic was proposed to address this issue.

The new technique utilises the earlier experience of experts and the efficiency of fast algorithms for feature ranking as well as the optimisation capabilities of a GA. A three-step hybrid algorithm that includes search space reduction, refining of the reduced feature subset, and incremental stages was designed. Search space reduction can quickly remove most of the irrelevant features. Refining of the reduced feature subset then further examines the reduced feature set. An incremental stage additionally improves the model's performance. To estimate its efficiency, this hybrid algorithm was applied to two real-world credit datasets: a Croatian and German dataset.

The obtained average accuracy rate was compared by the parametric paired *t*-Test and the nonparametric Wilcoxon matched-pairs signed rank test to that obtained by the GA-NN technique. The results of the statistical tests indicate that the available data support the hypothesis of the research at a 1% significance level.

Because the problems of feature selection are ubiquitous in data mining activities, a topic of future research can be the application of the presented algorithm to datasets in other areas. In this way, depending on the characteristics of the data, different fast filter techniques can be used in the search space reduction. Finally, from the comparison of the results on the German dataset, it is clear that this heuristic outperforms the results published in the literature. This finding additionally supports the hypothesis that the search space reduction and the incremental stage, combined with the effects of the GA initial population generating strategy, results in the enhancement of the classifier performance. This enhancement is large enough to be scientifically and practically interesting. Hence, the HGA-NN heuristic can be regarded as a promising addition to existing data mining techniques.

References

Aha, DW., & Bankert, R. L. (1996). *A comparative evaluation of sequential feature selection algorithms*. *Learning from data*. New York: Springer (pp. 199–206). New York: Springer.

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222*(1), 168–178.

Bache, K., & Lichman, M. (2013). *UCI machine learning repository*. Irvine, CA: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>.

Bäck, T., Fogel, D. B., & Michalewicz, Z. (1997). *Handbook of evolutionary computation*. IOP Publishing.

BIS. Basel III: a global regulatory framework for more resilient banks and banking systems. (2011). Basel Committee on Banking Supervision, Bank for International Settlements, Basel. ISBN print: 92-9131-859-0. <http://www.bis.org/publ/bcbs189.pdf>.

Danenas, P., Garsva, G., & Gudas, S. (2011). Credit risk evaluation model development using support vector based classifiers. *Procedia Computer Science, 4*, 1699–1707.

Finlay, S. (2011). Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research, 210*(2), 368–378.

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning, 3*, 95–99.

Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications, 33*, 847–856.

Jin, C., Jin, S. W., & Qin, L. N. (2012). Attribute selection method based on a hybrid BPNN and PSO algorithms. *Applied Soft Computing, 12*, 2147–2155.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*, 2767–2787.

Khashman, A. (2010). Neural networks for credit risk evaluation: investigation of different neural models and learning schemes. *Expert Systems with Applications, 37*, 6233–6239.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the ninth international workshop on machine learning* (pp. 249–256). Morgan Kaufmann Publishers Inc.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.

Li, X., Xiao, N., Claramunt, C., & Lin, H. (2011). Initialization strategies to enhancing the performance of genetic algorithms for the *p*-median problem. *Computers & Industrial Engineering, 61*, 1024–1034.

Maaranen, H., Miettinen, K., & Mäkelä, M. M. (2004). Quasi-random initial population for genetic algorithms. *Computers and Mathematics with Applications, 47*, 1885–1895.

Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega, 31*(2), 83–96.

McCall, J. (2005). Genetic algorithms for modeling and optimization. *Journal of Computational and Applied Mathematics, 184*, 205–222.

Michalewicz, Z. (1998). *Genetic algorithms + data structures = evolution programs*. Springer.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press/Addison-Wesley.

Myers, J. L., & Well, A. (2003). *Research design and statistical analysis*. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, Inc..

Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications, 39*, 12605–12617.

Pezzella, F., Morganti, G., & Ciaschetti, G. (2007). A genetic algorithm for the flexible job-shop scheduling problem. *Computers and Operations Research, 35*(10), 3202–3212.

Piramuthu, S. (2006). On preprocessing data for financial credit risk evaluation. *Expert Systems with Applications, 30*, 489–497.

Renner, G., & Ekárt, A. (2003). Genetic algorithms in computer aided design. *Computer-Aided Design, 35*, 709–726.

Šušteršic, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications, 36*, 4736–4744.

Tsai, M.-C., Lin, S.-P., Cheng, C.-C., & Lin, Y.-P. (2009). The consumer loan default predicting model – an application of DEA–DA and neural network. *Expert Systems with Applications, 36*, 11682–11690.

Wang, S. J., Mathew, A., Chen, Y., Xi, L. F., Ma, L., & Lee, J. (2009). Empirical analysis of support vector machine ensemble classifiers. *Expert Systems with Applications, 36*(3), 6466–6476.

Yang, W., Li, D., & Zhu, L. (2011). An improved genetic algorithm for optimal feature subset selection from multi-character feature set. *Expert Systems with Applications, 38*, 2733–2740.

Zhang, G., Gao, L., & Shi, Y. (2011). An effective genetic algorithm for the flexible job-shop scheduling problem. *Expert Systems with Applications, 38*, 3563–3573.