

AIRMS: A risk management tool using machine learning

Spyros K. Chandrinou^a, Georgios Sakkas^b, Nikos D. Lagaros^{a,*}

^a Institute of Structural Analysis & Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, 9, Heroon Polytechniou Str., Zografou Campus, GR-15780 Athens, Greece

^b School of Electrical & Computer Engineering, National Technical University of Athens, 9, Heroon Polytechniou Str., Zografou Campus, GR-15780 Athens, Greece



ARTICLE INFO

Article history:

Received 19 July 2017

Revised 22 March 2018

Accepted 23 March 2018

Available online 27 March 2018

Keywords:

Deep learning

Decision trees

Trading strategies

Risk management

Modified Renko bars

Profitable portfolio

ABSTRACT

In this work, an innovative approach to exploit the true potentials of machine learning (ML) by the financial industry is presented; using ML technology not as a source of investment ideas but as a consultant for trading decisions. In particular, the artificial intelligent risk management system (AIRMS) is presented that introduces one of the first efforts in the literature to utilize supervised ML as a risk management tool. Two AIRMSs systems are developed based on two well-known ML algorithms, i.e. artificial neural networks and decision trees. These two systems are applied into the five major currency pairs (FOREX) using signals obtained from an existing technical breakout trading strategy introduced in previous study by the authors, covering a seven-year period (2010–2016). Technical indicators and times series of the past entry points feed AIRMS in order to classify produced signals from the trading strategy into two classes: profitable and not. Constructing new portfolios using signals classified only as profitable resulted in an increased profit of more than 50% compared to the original ones. In this work, technical improvements are also proposed on the application of ML algorithms to financial data related to evaluation metrics and smoothing inputs. The obtained results revealed that the two AIRMSs can achieve impressive improvements to the performance of already profitable portfolios and proved that using ML to build risk management tools is very promising.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The first priority of investors is to construct a portfolio that could provide them with the highest possible return of the initial capital, especially now that the interests and safety that banks, used to offer, no longer exist. Selecting a portfolio is not a simple but rather a complicated procedure since there are thousands of portfolios in the market that an investor could choose. Apart from the arithmetic average return of the portfolio, the standard deviation of its returns is a major importance factor that an investor should consider in order to access the quality of a portfolio. The standard deviation is a crucial parameter, which affects significantly an important metric of a portfolio such as Sharpe ratio. If two portfolios have the same arithmetic average, the portfolio with the highest standard deviation will result in a lower Sharpe ratio. The highest standard deviation a portfolio has, the riskier it is. As Markowitz (1952) assumed in his *Modern Portfolio Theory*, investors are risk-averse, meaning that between two portfolios having the

same expected return investors would choose the one with the lower risk. An investor would accept increased risk, only if counterbalanced by higher expected return.

During the last two decades, a trending field of study is the application of artificial intelligent (AI) techniques in finance trying to generate profits mainly by forecasting the future movement of the market. Kimoto, Asakawa, Yoda, and Takeoka (1990) used back-propagation neural networks to predict the price of Tokyo Stocks and then determining buying and selling points. A noteworthy result of this study is that it succeeded prediction with very high accuracy. Refenes, Zapranis, and Francis (1994) used neural networks to develop a model based on arbitrage pricing theory in order to forecast the stock ranking. The application of the neural networks achieved better performance compared to the classical statistical techniques in terms of accuracy. Chen, Leung, and Daouk (2003) modeled the market index of Taiwan Stock Exchange trying to predict its future direction. They concluded that the investment strategies based on the probabilistic neural networks succeeded to outperform other common strategies such as the buy-and-hold one. Wu, Lin, and Lin (2006) used the decision tree algorithm to trade stocks based on filter rules. Applying this approach to Taiwan and NASDAQ stock markets, the authors succeeded better performance compared to the results of the filter

* Corresponding author.

E-mail addresses: chandrinou.spyros@gmail.com (S.K. Chandrinou), george.p.sakkas@gmail.com (G. Sakkas), nlagaros@central.ntua.gr (N.D. Lagaros).

rules approach itself. Choudhry and Garg (2008) provided a combined approach of support vector machine (SVM) with genetic algorithms (GA) for predicting stocks' future direction. GA was used for selecting the input features, which contained technical indicators and correlation metrics within the stocks. Also, their study proved that the combination of GA and SVM outperformed the application of SVM only. Chang, Liu, Lin, Fan, and Ng (2009) created a three-stage system based on dynamic time windows, case-based reasoning and a back-propagation neural network approach for stock trading prediction. Their attempt succeeded to result in high rates of returns for nine stocks with different kind of trends (upward, downward, steady). Teixeira and Oliveira (2010) conducted a research developing a trading system based on technical analysis (moving averages and relative strength indexes) and nearest neighbours' classification. Their analysis was applied to 15 stocks selected out of the 450 companies listed in Sao Paulo Stock Exchange (Bovespa); however, the 15 companies were chosen without any specific criteria making the evaluation of the results based on k-NN statistically insignificant. Furthermore, the buy-and-hold returns of stocks benchmark do not provide reliable evidence concerning the ability of k-NN to generate alpha, since the comparison of their returns is meaningless when these stocks do not outperform the Bovespa index. Moreover, the model is tested only to one direction trades (long) without evaluating its performance to short trades when this benchmark is used. Also, recently, Arevalo et al. (2017) proposed an automatic and dynamic trading rule based on the identification of flag patterns using several filters. Applying this system to DJIA index, they succeeded to outperform the previous results in this field.

Most of the existing studies such as Chang et al. (2009), Teixeira and Oliveira (2010), and Chen et al. (2003), which already have been referred above, depict similar technical deficiencies in the way that they try to approach the efficiency of applying an algorithm to financial data. They do not use deeply statistical analysis in order to conclude to more generalized and broad results. Furthermore, we argue that the usage of ML "black box" algorithms as the main trading tool is not efficient and applicable under real trading circumstances.

The novelty of our study relies on the use of machine learning (ML) in the investing field as a risk management tool. This work develops a systematic approach that can be used by the expert system with applications (ESWA) community as an expert advisor improving the performance of existing trading strategies and generating valuable alpha for the financial industry. In particular, the study will focus on the classification of the produced signals of a trading strategy into profitable and non-profitable ones by means of artificial neural networks (ANN) and decision trees (DT). Specifically, the two AI techniques will be applied on five currency pairs (GBP/USD, USD/JPY, EUR/USD, GBP/JPY and EUR/JPY) through a channel breakout trading strategy that was recently introduced by Chandrinós and Lagaros (2018). The aim is to improve the performance of the enhanced equally weighted portfolio (EWP2) and the enhanced Kelly criterion portfolio (KCP2) portfolios that were constructed by the authors in their recent study, focusing on the reduction of their standard deviation and the amelioration of the correlation between their total return and the portfolios' standard deviation. The sharpe ratio is the measure that will be used to approximate this relationship. The two above mentioned machine learning techniques are based on a training-testing-evaluation procedure that in the current study follows a dynamic sliding window approach. The size of the training windows is three years of currencies exchange rates performance data, while the size of the testing and evaluation period is equal to one year each. During the training-testing procedure, both AI techniques are calibrated based on several input parameters. The F1-score is the evaluating metric that is used for both AI techniques. The objective of the calibration

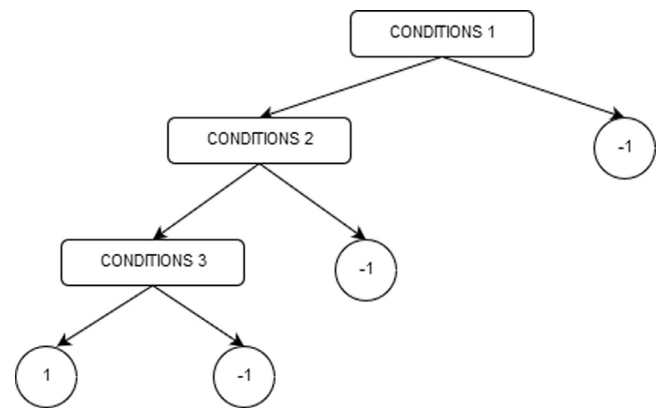


Fig. 1. Binary decision tree.

procedure is to predict the quality of the produced trading signals by the MRB channel strategy and avoid the signals that might lead to a losing trade. The signals that will be classified as profitable via the developed AI based risk management system (AIRMS) of each AI techniques are transformed to actual trades that will be used in order to construct the new AI-based equally weighted portfolios (EWP-DT and EWP-ANN) and the new AI-based Kelly criterion portfolios (KCP-DT and KCP-ANN) following the same principals as the original ones. Finally, based on the results obtained the AIRMS will be evaluated.

2. Machine learning background

In this work, two machine learning techniques are employed, namely decision trees and deep neural networks, and a short description for both of them is provided below. Machine learning models aim to "learn" the relationship between a set of input and output data, i.e. attempt to create the desired mapping between the inputs and the targets of a set that is composed by np input-target pairs $D = [x^{np}, t^{np}] \in R^{ni} \times R^{no}$. The first coordinate is a position in ni -dimensional space, corresponding to the input space, and the second coordinate is a position in no -dimensional space, corresponding to the desired or target space. The ML techniques are used in order to form the mapping $R^{ni} \rightarrow R^{no}$ between those two spaces.

2.1. Decision trees

Decision trees (DTs) represent a non-parametric supervised learning technique, often used for classification and regression purposes. The goal is to create a model able to predict the value of a target variable by learning simple decision rules inferred from the data features. There are several variants of decision trees such as ID3, C4.5, C5.0 and classification and regression trees (CART). An optimized version of CART is used in this work (see Fig. 1).

2.1.1. CART algorithmic description

Globally-optimal classification tree analysis (GO-CTA) (also called hierarchical optimal discriminant analysis) refers to the generalization of optimal discriminant analysis. It can be used to identify the statistical model that achieves the maximum accuracy for predicting the value of a categorical dependent variable for a dataset consisting of categorical and continuous variables. The resulted GO-CTA model is a non-orthogonal tree that combines categorical variables and cut points for continuous variables; yielding maximum prediction accuracy. Hierarchical optimal discriminant analysis may be considered as a generalization of Fisher's linear discriminant analysis. Optimal discriminant analysis is an alternative to the analysis of variance (ANOVA) and regression analysis,

which attempt to express one dependent variable as a linear combination of other features or measurements. However, ANOVA and regression analysis give a dependent variable that is a numerical variable, while hierarchical optimal discriminant analysis gives a dependent variable that is a class variable.

If a target is a classification outcome which taking a discrete value, for node m , representing a region R_m with N_m observations, let:

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k) \quad (1)$$

be the proportion of class k observations in node m .

CART as a classification method contains some important advantages compared to other statistical classification methods when applied to the real sector. The first one is that CART is a parametric-free method since it does not require specification of any functional form. Therefore, this algorithm will itself identify the most significant variables and eliminate non-significant ones. Besides, CART can easily handle outliers that usually have a negative effect on the results of some statistical models such as principal component analysis (PCA) and linear regression. The splitting algorithm of CART easily handles noisy data and the outliers are isolated into a separate node. This property is very important because financial data very often contain outliers due to significant financial events, like defaults. Another advantage of CART, which is important for the financial sector, is its flexibility and ability to adjust in time, i.e. the learning sample is consistently replaced with new observations. It means that CART has an important ability to adjust to the current situation in the market (Breiman, Friedman, Olshen, & Stone, 1984).

2.1.2. Parameters of DT

In each training-testing session, eight parameters of DT are calibrated in order to reach to the DT model that will give the highest evaluation metric. The *Criterion* denotes the function to measure the quality of a split. Gini impurity and Cross-Entropy are the two criterions that are used. *Splitter* represents the strategy that is used to choose the split at each branch of the DT (referred also as node). The supported strategies are “best” to choose the best split and “random” to choose the best random split. *Max features* are the maximum number of features to consider when looking for the best split. *Min samples split* is the minimum number of samples required to split an internal node. *Max depth* is the maximum depth of the tree. It is possible that nodes are expanded until all leaves are pure or until all leaves contain less than *min samples split* samples. *Min samples leaf* is the minimum number of samples required to be at a leaf node. *Class weights* represent the weight coefficients that are associated with classes. It is possible that all classes are supposed to have weight one. *Min impurity split* is the threshold for early stopping in tree growth. A node will split if its impurity is above the threshold, otherwise, it is a leaf.

2.2. Deep neural networks

ANNs represent a biologically inspired computational machine learning model; they are composed of elements known as artificial neurons that perform in a manner analogous to the elementary functions of a biological neuron. ANN's organization is inspired heavily by the anatomy of the brain and they exhibit a surprising number of the brain's characteristics. Typically, the neurons are arranged into layers and the signals travel from the first layer (labeled as input) through hidden ones (if any) to the last one (labeled as output) (see Fig. 2).

2.2.1. ANN algorithmic description

The objective of ANNs is to solve problems in the same way that a human brain would, like learning from experiences; thus,

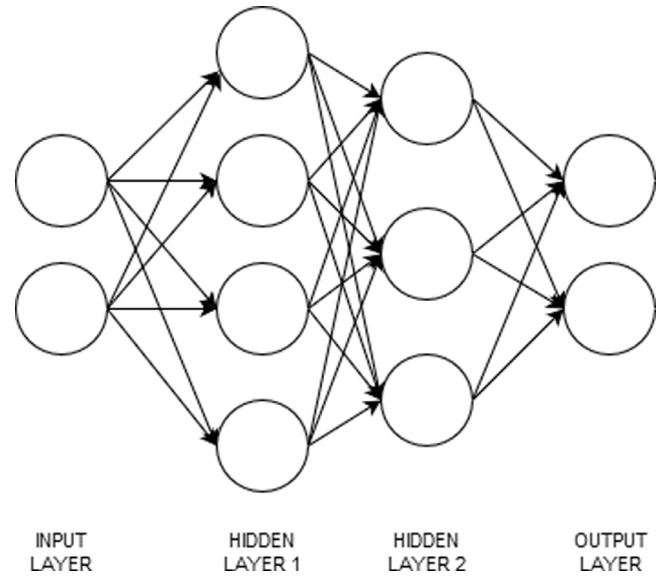


Fig. 2. A four-layer feedforward neural network.

ANNs “understand” past data in order to use them for future predictions. Pattern recognition is one of the most common application areas of neural networks. Pattern recognition models are in many cases calibrated over a pool of labeled data called “training” data (supervised learning) when no labeled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning). This study focuses on supervised learning through classification. ANNs process records one at a time, and “learn” by comparing their classification of the record (which, at the outset, is largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record are fed back into the network and used to modify the network's algorithm the second time around, and so on for many iterations.

A Neural Network architecture A is structured by layers (an input layer, one or more hidden ones and an output one), a number of units in each layer and a type of activation function. If a set of weight parameters \mathbf{w} is assigned to the connections of the network, a mapping $\mathbf{y}(\mathbf{x}^{np}; \mathbf{w}, A)$ is defined between the input vector \mathbf{x}^{np} and the output vector \mathbf{y} . The quality of this mapping is measured using the following error function:

$$E_D(D|\mathbf{w}, A) = \sum_m \frac{1}{2} [y(\mathbf{x}^{np}; \mathbf{w}, A) - t^m]^2 \quad (2)$$

A learning algorithm tries to determine the weight parameters \mathbf{w} in order to minimize the error function E_D . Iterative minimization algorithms are therefore used to obtain the optimum weight parameters \mathbf{w} . For example, a stochastic gradient descent algorithm can be used for the solution of the minimization problem, which, given a random starting vector of weight parameters $\mathbf{w}^{(0)}$, updates the weights by $\Delta \mathbf{w}^{(t)}$ at each iteration, where $\Delta \mathbf{w}^{(t)}$ is proportionate to the direction (gradient) of the minimum and a given learning rate.

In this work, the feedforward neural network (FFNN) is the type of ANN that is used. The special characteristic of this neural network is that the neurons between the units do not form a cycle, meaning that the information moves in only one direction, forward, from the input nodes, through the hidden nodes, and to the output nodes. Due to the non-linear nature of the problem that this study tries to solve, multilayer FFNNs are applied, which are also called deep neural networks (DNNs). As Hornik, Stinchcombe, and White (1989) supported and it is also mathematical proved,

DNNs are considered universal approximators. This term means that they are capable of approximating any measurable function to any desired degree of accuracy. There are no theoretical limitations of the success of DNNs. So, any lack of success is related to inadequate learning, an insufficient number of hidden units or the lack of deterministic relationship between input and target.

2.2.2. The hyperparameters of the ANN

The type and the performance of ANN are affected by its parameters, i.e. number of *hidden layers* and their *processing elements*. In an effort to identify the optimum number of these parameters general tests were conducted for two and three hidden layers. The architecture having two hidden layers provided consistently better results. *Processing elements per hidden layer* is an important decision for the correct structure of a neural network. In this work, the number of processing elements in each hidden layer is determined by an optimization through the train data in order to succeed the higher results based on the used evaluation metric. *Early stopping* is a form of regularization used to avoid over-fitting when training a learner with an iterative method, such as gradient descent. Such methods update the learner in every iteration so as to make it better fit the training data. Up to a point, this improves the learner's performance on data outside of the training set. Past that point, however, improving the learner's fit to the training data comes at the expense of increased generalization error. Early stopping rules provide guidance as to how many iterations can be run before the learner begins to over-fit. Early stopping rules have been employed in many different machine learning methods, with varying amounts of theoretical foundation. In our calculations, the early stopping parameter is defined as six maximum validation failures. *Iterative Learning Process* is a procedure in which data cases are presented to the network one at a time, and the weights associated with the input values are adjusted each time. After all cases are presented, the process often starts over again. During this learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of input samples. This developed risk management system uses the gradient descent with momentum and adaptive learning rate back-propagation as an iterative learning process. Gradient descent is a first-order iterative optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point. An extension of this gradient descent algorithm is the momentum method that emphasizes on reducing the risk of getting stuck in a local minimum, as well as speeds up the convergence considerably in cases where the process would otherwise zig-zag heavily. The term "back-propagation" means the most common training procedure for feedforward neural networks that consists in an iterative optimization of a so-called error function representing a measure of the performance of the network. As Li, Fu, Li, and Zhang (2009) introduced in order to avoid oscillation inside the network, such as alternating connection weights, and to improve the rate of convergence, there are refinements of the back-propagation algorithm that use an adaptive learning rate.

2.3. Evaluation metric

As it can be seen, each ML algorithm consists of a variety of parameters. For different values of each parameter of the ML algorithms, a different prediction model is formed. In order to determine the most proper model for our purposes, a suitable evaluation metric should be used. Evaluation metrics are the criteria used in order to evaluate the ML algorithms (ANN and DT) success and ameliorate their predicting ability. The algorithm, based on a metric, attempts to 'understand' the meaning of its purpose. The major objective of the study is to increase the total return and to

reduce the standard deviation of portfolios. Originally, traditional metrics, like the accuracy between predictions and targets, were used but were found to be not so appropriate for our task. These metrics would consider the two classes as equals and each signal as equally weighted, and that is not our case. Therefore, this would lead to non-standard profits; i.e. similar accuracies for two different models would lead to completely different profits. Obviously, there was a necessity to use a different metric, one that would not have the aforementioned drawbacks. For these reasons the metric that was used is the F1-score, which is commonly used in statistical analysis of binary classification. A high F1-score means high precision (p) and recall (r) for the given model, as it is the harmonic mean if those two. Thus having a high F1-score means that the model's predictions about the winning trades are usually correct and the model can predict almost all of them. Consequently, having a good F1-score guarantees that the model predicts correctly most of the winning trades and they are not polluted by many losing ones. Therefore, the overall profit will not drop down as happened in the case that accuracy was used as a metric, and the aforementioned drawbacks vanish.

$$F1 = 2 \cdot \left[p \cdot \frac{r}{(p+r)} \right] \quad (3)$$

2.4. Training, testing, and validation sets

In this work, the approach in order to evaluate both AI techniques is a walk-forward testing routine also known as sliding or moving window testing. Popular in evaluating commodity trading systems, walk-forward testing involves dividing the data into a series of overlapping training-testing-validation sets. This approach attempts to simulate real-life trading and tests the robustness of the model through frequent retraining on a large out-of-sample data set. In walk-forward testing, the size of the validation set leads the retraining frequency of the machine learning algorithm. Frequent retraining is more time consuming but allows the decision tree to adapt more quickly and accurately to changing market conditions. Furthermore, the training and testing sets are scaled together, since the purpose of the testing set is to determine the ability of the AI method to generalize.

In previous study, Chandrinis and Lagaros (2018) employed data from five currencies pairs for the period 2006–2016. In this study, the data of the first four-year period (2006–2009) were used to optimize the trading system's parameters, while the application of the optimized system corresponds to the 2010–2016 period. Based on the walk-forward testing routine, the size of the training window is a three-year period, the size of the testing window is the following year after the training period and the evaluation set is the next year after the testing year. This model slides through the pass of every year. Specifically, the signals that are produced by the MRB channel strategy for each currency pair during the three-year period of 2006–2008 are used as training data for the ANNs and DTs. The signals from 2009 represent the testing data of the model, while the evaluation data will be the produced signals during 2010. So, the developed system will attempt to predict which signals of 2010 will be profitable and which will be not. After the end of 2010, the whole system moves forward by one year meaning that the new training set is the 2007–2009 period, the new testing period is the year of 2010 and the evaluation period is 2011 (see Fig. 3).

The return of each year for each pair based on each machine learning algorithm is determined as the sum of the returns of all the trades that the algorithm classified their signal as profitable. The signals that were categorized as non-profitable will be ignored.

Table 1 illustrates the way each AIRMS manages the produced signals. These signals constitute a subset of the signals that produced during 2010 in GBP/USD. The mark "1" means that the signal

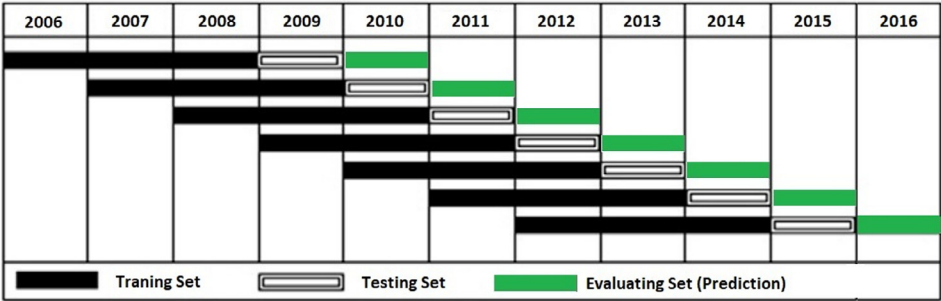


Fig. 3. Walk-forward sliding windows testing routine.

Table 1
Predictions of AIRMS-ANN and AIRMS-DT.

No. of signals	Profit/loss (pips)	Predictions			Results (pips)		
		without AIRMS	AIRMS-ANN	AIRMS-DT	without AIRMS	AIRMS-ANN	AIRMS-DT
1	(28.40)	1	1	−1	(28.40)	(28.40)	0.00
2	(48.00)	1	1	−1	(48.00)	(48.00)	0.00
3	143.40	1	1	1	143.40	143.40	143.40
4	53.60	1	1	1	53.60	53.60	53.60
5	153.50	1	1	1	153.50	153.50	153.50
6	(15.60)	1	1	−1	(15.60)	(15.60)	0.00
7	(52.10)	1	1	1	(52.10)	(52.10)	(52.10)
8	(74.40)	1	1	−1	(74.40)	(74.40)	0.00
9	324.00	1	1	1	324.00	324.00	324.00
10	(71.90)	1	1	1	(71.90)	(71.90)	(71.90)
11	(36.00)	1	1	−1	(36.00)	(36.00)	0.00
12	181.90	1	1	1	181.90	181.90	181.90
13	60.00	1	1	−1	60.00	60.00	0.00
14	231.80	1	1	1	231.80	231.80	231.80
15	(96.00)	1	1	−1	(96.00)	(96.00)	0.00
16	24.00	1	1	−1	24.00	24.00	0.00
17	(85.00)	1	1	−1	(85.00)	(85.00)	0.00
18	20.90	1	1	−1	20.90	20.90	0.00
19	20.70	1	−1	1	20.70	0.00	20.70
20	(42.70)	1	1	1	(42.70)	(42.70)	(42.70)
21	(12.00)	1	1	−1	(12.00)	(12.00)	0.00
22	28.70	1	1	1	28.70	28.70	28.70
23	(48.00)	1	−1	−1	(48.00)	0.00	0.00
24	58.40	1	−1	1	58.40	0.00	58.40
25	(60.00)	1	1	−1	(60.00)	(60.00)	0.00
26	(17.50)	1	1	−1	(17.50)	(17.50)	0.00
27	(46.80)	1	−1	1	(46.80)	0.00	(46.80)
28	(75.20)	1	1	−1	(75.20)	(75.20)	0.00
29	(81.40)	1	−1	1	(81.40)	0.00	(81.40)
30	61.80	1	1	1	61.80	61.80	61.80
Total return (pips)					471.70	568.80	962.90

would be a profitable one according to the corresponding AIRMS so it is formed to a trade. The mark “−1” means that the signal would be a non-profitable one so it is ignored. Based on this logic the total return of the thirty trades is formed for each AIRMS. The column “Without AIRMS” represents the original MRB trading strategy where all the signals become trades.

3. Features

The features that are used as inputs to the ML models aim to describe the class that every specific object belongs to. In this work, in order to help investors to follow the signals that are most likely to generate profit, several input features are combined, aiming to predict when a produced trading signal will conclude to a winning or a losing trade. In total, the number of the input features considered in this study is equal to 20. Most of these features (14 out of the 20 ones) correspond to technical indicators that are calculated based on the prices of the MRB chart. The values of these indicators are defined at the specific time that a signal is produced.

Technical indicators refer to mathematical expressions that are used to predict the financial market direction. These mathematical formulas are based mainly on historic price (open, close, high and low) and volume. Generally, indicators overlay on price chart data to indicate where the price is directed, or whether the price is in an “overbought” or “oversold” state.

The features based on technical indicators that are used in this study are: (i) Cross of simple moving average of the last 200 bars and simple average of the last 100 bars (one feature), (ii) Cross of simple moving average of the last 50 bars and simple average of the last 25 bars (one feature), (iii) Relative strength index of the last 14 MRB (one feature), (iv) Acceleration between 12 last bars (one feature), (v) Moving average convergence/divergence (MACD) (one feature), (vi) Stochastic oscillator (one features), (vii) Momentum of the 12 last MRB (one features), (viii) Bollinger band (two features; upper, lower and middle band), (ix) Weighted close price (one feature), (x) Price rate of change of the close price of the last 12 bars (one feature), (xi) High and low channel of the last Y MRB (two features). Where the Y has been computed for every pair in a

recent study by Chandrinou and Lagaros (2018). These features are described below:

Simple moving average (SMA) is an arithmetic moving average calculated by adding the closing price of the security for a number of bars and then dividing by the number of bars.

Relative strength index (RSI) is a momentum indicator that compares the magnitude of recent profits and losses over a specific number of bars to calculate speed and change of price movement of a security. It is mainly used to identify overbought or oversold state in the trading of an asset. It has a range between 0 and 100. Generally, when the RSI is under the level 30 is oversold and when is above the level 70 is overbought.

$$RSI = 100 - \left[\frac{100}{(1 + RS)} \right] \quad (4)$$

where *RS* denotes the Relative Strength that is calculated as the ratio of average gain / average loss.

Moving average convergence divergence (MACD) is a trend-following momentum indicator that presents the relationship between two moving averages of prices. The *MACD* is computed by subtracting the 26-bar *exponential moving average (EMA)* from the 12-bar *EMA*. A nine-day *EMA* of the *MACD* called the “signal line”, is then plotted on top of the *MACD*, functioning as a trigger for buy and sell signals.

$$\begin{aligned} MACD &= EMAofC(N = 12) - EMAofC(N = 26) \\ SignalLine &= EMAofMACD(N = 9) \end{aligned} \quad (5)$$

where

$$EMAofX(N) = X_n \frac{2}{(N+1)} + X_{n-1} \left[1 - \frac{2}{(N+1)} \right] \quad (6)$$

Momentum is a simple technical indicator that shows the difference between recent closing price and the close *N* bars ago.

$$Momentum = |C_{current} - C_{current-N}| \quad (7)$$

Bollinger band, developed by famous technical trader John Bollinger, is a volatility channel which its levels is determined two standard deviations away from a simple moving average of *N* bars. Where the *middle band* is equal to *SMAofC(N)* and

$$\begin{aligned} UpperBand &= MiddleBand + 2 \cdot stDev(lastNbars) \\ LowerBand &= MiddleBand - 2 \cdot stDev(lastNbars) \end{aligned} \quad (8)$$

The *high and low channel* is a moving average indicator that tracks a security's highest highs and lowest lows over a set number of bars. Where the *high channel* is equal to the maximum of the *N* last bars high and *low channel* is equal to the minimum of the *N* last bars low.

Stochastic oscillator is a momentum indicator that compares the closing price of a security to the range of its prices over a certain number of bars. The sensitivity of the oscillator to market movements is reducible by taking a moving average of the result.

$$\begin{aligned} K &= 100 \frac{C - L_{10}}{H_{10} - L_{10}} \\ D &= SMAofK(N = 3) \end{aligned} \quad (9)$$

where L_{10} is the minimum of the 10 last bars low and H_{10} is the maximum of the 10 last bars high.

Price of change (PoC) calculates the price rate of change from the closing price. The price rate of change is calculated between the current closing price and the closing price *N* bars ago.

$$PoC = \frac{(C_{current} - C_{current-N})}{C_{current-N}} \quad (10)$$

Weighted close price (WCP) attempts to normalize the movement of the price. So it is calculated as the average of twice the closing price plus the high and low prices.

$$WCP = \frac{(2 \cdot C + H + L)}{4} \quad (11)$$

Table 2

Number of selected features via k-best algorithm for calibrating the AIRMS-DT.

Year period	Number of selected features for AIRMS-DT				
	GBP/USD	USD/JPY	EUR/USD	GBP/JPY	EUR/JPY
2010	13	19	16	6	5
2011	12	9	7	8	13
2012	7	17	12	11	16
2013	12	10	14	5	10
2014	7	8	18	11	17
2015	15	5	10	18	5
2016	5	16	10	20	6

Acceleration (AC) is the difference of two momentums separated by *N* bars.

$$AC = Momentum_{current} - Momentum_{current-N} \quad (12)$$

The rest six features are based on some specific characteristics of the way the entry signals of the strategy are formed. Thus, five of these six features contain the last five entry prices where a signal was produced. The idea behind this selection is for the two algorithms to create potential support and resistance levels, which would help them to identify more accurately the quality of the signals. The last feature is focused on the difference between the *MRB* close price and the true price of the currency at the time of the signal. The absolute value of this difference is calculated. It is reasonable that the higher this value is, the smaller the probability of the signal to end into a winning trade will be.

4. Numerical tests

The numerical investigation is composed by the implementation of the walk-forward testing routine to both decision trees and artificial neural networks models adopted for extracting the returns for each currency, the reconstruction part of the portfolios and the comparison to the original ones. A set of four-year signals is used for training and testing (75% of the set is used for training and the rest 25% for testing) while the signals of the next year are used for the evaluation part. This mechanism slides forward at the end of each year by one year (see Fig. 2).

4.1. Calibration of decision trees

Decision trees have value even with little hard data and they also use a white box model. Based on these characteristics, another parameter that is used in order for the DT algorithm to be calibrated in each training-testing session is the *k* (smaller or equal to twenty) best features from the total 20 features. All possible combinations of the measures given in Table 2 are assessed with respect to the median predicting errors on the testing set. If *n* is the total number of features then the total number of combinations is equal to:

$$\sum_{v=5}^n \frac{n!}{v!(n-v)!} \quad (13)$$

where *n* is equal to 20 in this study. Tests which were conducted to this kind of data showed that the DT algorithm response better to a fewer number of features. The *k*-best model is applied each time to the corresponding training-testing data testing all the possible combination of the feature set contributing to the amelioration of the performance of the DT algorithm. The criterion that the *k*-best algorithm uses in order to select the *k*, each time, best features is the *F*-value of ANOVA. *F*-test is any statistical test in which the test statistic has an *F*-distribution under the null hypothesis. It is most

Table 3
Currency returns based on AIRMS-DT.

Years	Returns via AIRMS-DT (%)				
	GBPUSD	USDJPY	EURUSD	GBPJPY	EURJPY
2010	16.21	(47.17)	104.24	35.43	102.21
2011	(25.53)	0.90	35.97	19.22	77.18
2012	37.32	45.74	(10.50)	36.46	52.19
2013	(37.87)	33.47	1.54	3.36	15.54
2014	3.54	45.53	59.02	68.19	59.07
2015	20.45	12.50	15.23	(42.81)	(48.82)
2016	97.30	32.21	(11.22)	120.69	16.22
Total return (%)	111.42	123.18	194.28	240.54	273.59

Table 4
Currency returns without any AIRMS.

Years	Returns via MRB channel strategy				
	GBPUSD	USDJPY	EURUSD	GBPJPY	EURJPY
2010	(2.97)	(51.89)	104.24	35.43	77.89
2011	(29.98)	(22.72)	(2.41)	17.08	77.18
2012	17.00	39.96	(13.57)	36.46	62.35
2013	(31.20)	17.42	0.11	10.83	(20.43)
2014	(11.64)	42.34	69.89	68.19	34.60
2015	12.51	9.73	15.23	(29.14)	(49.95)
2016	97.30	40.02	(17.70)	78.74	27.10
Total return (%)	51.02	74.86	155.79	217.59	208.74

often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. In one-way ANOVA, the F-statistic is calculated as the ratio: *F*= *variation between sample means* / *variation within the samples*. The features that were selected as inputs to the DT of each currency pair for every year are shown in Table 2.

Yearly results obtained by the implementation of the DT based AIRMS for each currency are depicted in Table 3. For each FX pair, the results for the year 2010 were produced using the DT model, calibrated based on the data of the previous four years (i.e. 2006–2009). Hence, by means of the training and testing procedures, using the data of the first three years (75%) those of the fourth one (25%), respectively, it was succeeded to identify the best value of the models’ parameters which provided the highest F1-score. Using the calibrated model, the class of the signals of the year 2010 is predicted. Based on the walk-forward testing routine the same procedure is followed to calculate the returns of the remaining six years. More specifically, in order to predict the return of the year 2011, the period of years 2007–2010 is employed as training-testing dataset up to the period of years 2012–2015 is used in order to calibrate appropriately the DT model able to predict which signals of 2016 will be transformed to an actual trade and which not. The sum of the outcome of that trades will lead to the return of 2016. In order to assess the performance of the DT based AIRMS, the produced returns of the system need to be compared to the original ones of the MRB channel strategy. The performance of AI-assisted MRB trading strategy and the conventional one is provided in Tables 3 and 4, respectively. In the following part of the session, a comparative presentation for each currency exchange rate is provided.

In the case of GBPUSD exchange rate, for six out of the seven years, the returns produced by the DT based AIRMS are higher compared to those achieved by the MRB strategy. Specifically, the largest increase related to the original results was achieved during the year 2010. The AIRMS succeeded to improve the return for the year 2010 by almost 645%, i.e. from the value of –2.97 to 16.21%. As it can be seen, AIRMS-DT provided more stable and profitable returns through the period of the 7 years reducing the fluctuations

of the MRB strategy. Also, the total return achieved for the AIRMS is equal to 111.42% that in comparison to the original one of 51.02% corresponds to a significant improvement of 118%.

In the case of USDJPY exchange rate, for six out of the seven years, the returns produced by the DT based AIRMS are also higher compared to those achieved by the MRB strategy. Specifically, the higher increase related to the original results was achieved during the year 2011. The AIRMS succeeded to improve the return for the year 2011 by about 104%, i.e. from the value of –22.72 to 0.9%. It can be observed that AIRMS-DT succeeded to reduce the standard deviation of the returns of USDJPY between 2010 and 2016 by 10%. The total return achieved for the AIRMS is equal to 123.18% that in comparison to the original return of 74.86% corresponds to a significant improvement of 64.55%. Concluding

Accordingly, in the case of EURUSD exchange rate, for six out of the seven years, the returns produced by the DT based AIRMS are higher compared to those achieved by the MRB strategy. Specifically, the largest increase related to the original results was achieved during the year 2011. The AIRMS succeeded to improve the return for the year 2011 by approximately the extreme value of 1500%, i.e. AIRMS succeeded to alter a negative return value of –2.41% to a positive one of 35.97%. Related to the standard deviation of the returns of EURUSD, AIRMS-DT achieved a decrease of 10%. Furthermore, the total return achieved for the AIRMS is equal to 194.28% that in comparison to the original return of 155.79% corresponds to a significant improvement of 24.70%.

In the case of GBPJPY exchange rate, for five out of the seven years, the returns produced by the DT based AIRMS are higher or equal compared to those achieved by the MRB strategy. Specifically, the largest increase related to the original results was during the year 2016. The AIRMS succeeded to improve the return of 2016 by about 53.28%, i.e. AIRMS achieved rise, even more, the already high return of 78.74–120.69%. The total return succeeded for the AIRMS is equal to 240.54% in comparison to the original return of 217.59%, corresponds to an increase of 10.55%.

In the case of EURJPY exchange rate, for five out of the seven years, the returns produced by the DT based AIRMS are higher compared to those achieved by the MRB strategy. Specifically, the largest increase in the original results was during the year 2013. The AIRMS succeeded to improve the return of 2013 by about 176%, i.e. AIRMS achieved to transform a negative return of –20.43% into to a positive one of 15.54%. AIRMS-DT succeeded to reduce the fluctuations between positive and negative returns, even more, providing a more steady environment for the investor. The total return achieved for the AIRMS is equal to 273.59% in comparison to the original return of 208.74%, corresponds to an increase of 31%.

A meaningful observation is that the developed risk management system based on DT achieved its purpose; to improve most of the negative returns, which produced the MRB channel strategy and to reduce the standard deviation of each individual currency pair. Hence, AIRMS-DT prevented the investors from the significant drawdowns of the original trading strategy. Another important observation is that it succeeded to ameliorate the total return for all currency pairs. These observations prove that the DT based AIRMS provides a stable capability to predict the class of the produced signals in such a way in order to generate constantly high profits and to decrease fluctuations.

4.2. Re-construction of EWP2 and KCP2 based on AIRMS-DT

In the previous section the returns for each currency pair, that were produced by the DT based AIRMS, were presented. In this section, the two portfolios constructed in previous study by Chandrinos and Lagaros (2018) are reconstructed. The first one is a simple average portfolio where all pairs have the same weight coefficient. The second portfolio is rather more advanced and differ-

Table 5

Allocation of the five currencies in EWP-DT.

Year period	Equally weighted portfolio based on AIRMS-DT (EWP-DT) (%)				
	GBP/USD	USD/JPY	EUR/USD	GBP/JPY	EUR/JPY
2010	20.00	20.00	20.00	20.00	20.00
2011	20.00	20.00	20.00	20.00	20.00
2012	20.00	20.00	20.00	20.00	20.00
2013	20.00	20.00	20.00	20.00	20.00
2014	20.00	20.00	20.00	20.00	20.00
2015	20.00	20.00	20.00	20.00	20.00
2016	20.00	20.00	20.00	20.00	20.00

Table 6

Equally Weighted Portfolio based on AIRMS-DT (EWP-DT).

Years	2010	2011	2012	2013	2014	2015	2016
EWP-DT Returns (%)	42.18	21.55	32.24	3.21	47.07	(8.69)	51.04

Table 7

Equally Weighted Portfolio without any AIRMS (EWP2).

Years	2010	2011	2012	2013	2014	2015	2016
EWP2 Returns (%)	32.54	7.83	28.44	(4.65)	40.68	(8.32)	45.09

ent weight coefficients are assigned to the various currency pairs. The weight coefficients of each currency pair for each year are determined by its performance during the previous year based on the Kelly criterion. The portfolios are formed based on exactly the same approach that was used for EWP2 and KCP2 presented in previous study by Chandrinós and Lagaros (2018). The two re-constructed portfolios will be denoted as EWP-DT and KCP-DT, respectively. In order to evaluate the effect of the DT based AIRMS on the performance of the two portfolios, EWP-DT and KCP-DT will be compared to the original ones (i.e. EWP2 and KCP2).

Table 5 shows the yearly proportions of each currency pair assigned to the EWP-DT portfolio. The allocation chosen is equal to 1/5 since EWP-DT is composed of five currency pairs. Table 6 contains the returns of EWP-DT portfolio, which is based on the AIRMS-DT, while Table 7 presents the returns of the initial portfolio EWP2 that was based on the original MRB channel strategy. Comparing the results of these two tables it can be observed that the re-constructed equally weighted portfolio succeeds to outperform the original one EWP2 for six out of the seven years. Specifically, the broadest increase related to the original results was achieved during the year 2011. EWP-DT portfolio succeeded to improve the return for the year 2011 by almost 175%, i.e. the return was improved from the value of 7.83–21.55%.

The improvements that the developed DT based risk management system achieved on the performance of EWP2 are more clearly presented when comparing the evaluation metrics of the two equally weighted portfolios. These metrics can be seen in Tables 8 and 9, which contain the total return, the compound growth, the arithmetic average, geometric return, standard deviation and the Sharpe ratio of each portfolio corresponding. Notably, EWP-DT portfolio attained a total return of 188.60% improving the total return of the original equally weighted portfolio by approximate 33%. Regarding compound growth, the re-constructed equally weighted portfolio reported a return equal to 478.43% that com-

pared to the EWP2's compound growth is increased by 46%. This means that an investor who invests \$10,000 in 2010 at EWP-DT, he would have \$47,800 by the end of 2016. Most importantly, the AIRMS-DT improved the two most significant metrics related to risk management field of the new portfolio; the standard deviation and the sharpe ratio. The standard deviation of the new average portfolio decreased by 3.20%, from 21.77 to 21.07%, making the EWP-DT less fluctuated. Also, the application of the AIRMS-DT to the EWP2 had as result to raise significantly the sharpe ratio by 42.43%, i.e. from 0.81 to 1.16. The huge improvement of the sharpe ratio is very crucial because investors use this metric to access the quality of a portfolio, combining the return and the standard deviation of a portfolio, measures risk-adjusted return.

Table 10 presents the yearly percentages of each currency pair assigned to the KCP-DT. This allocation during a year is based on the Kelly Criterion of each pair from the previous year. For this reason, the 2010's weight coefficients are equal for all the currencies due to the lack of data from 2009. Table 11 contains the returns of the KCP-DT portfolio, which is based on the AIRMS-DT, while Table 12 presents the returns of the KCP2 from the original MRB channel strategy. Comparing the results of these two tables, it can be observed that the re-constructed KC based portfolio succeeds to outperform the KCP2 for six out of seven years. Specifically, the bigger increase related to the original returns was during 2013. The KCP-DT succeeded to rocket the return for the year 2013 by about 303%, i.e. from 1.99 to 8.03%. This means that achieved to predict with great accuracy the most of non-profitable signals in order to avoid losses of those potential trades. Another significant amelioration which the KC-based portfolio succeeded by AIRMS-DT was during 2011, where the 18.30% return almost doubled it to 32.73% return.

These improvements that the developed DT based risk management system achieved on the performance of KCP2 are more clearly presented when comparing the evaluation metrics of the two KC based portfolios. These metrics can be seen in Tables 13 and 14 that contain the total return, the compound growth, the arithmetic average, geometric return, standard deviation and the Sharpe ratio for each portfolio corresponding. Notably, KCP-DT portfolio improved both the standard deviation and the sharpe ratio (the risk metrics). The standard deviation of the new KC portfolio is decreased by 3.5%, from 20.42% to 19.70%; and its sharpe ratio increased significantly by 30.55%, i.e. from 1.14 to 1.47.

Also, the new KC portfolio achieved a significantly high total return of 220.14% improving the total return of the original KC based portfolio by approximately 23.43%. Regarding the compound growth, the re-constructed portfolio based on KC achieved a return equal to 621.46% that compared to the KCP2's compound growth is increased by 38%. This means that an investment of \$10,000 in 2010 on the KCP-DT portfolio, a capital equal to \$62,140 will be returned in the year 2016.

The contentious improvements on the performance of the EWP2 and KCP2 portfolios, that the adjustment of the AIRMS-DT offered, underline that its ability to recognize the quality of the signal produced is significant. In both portfolios, the use of AIRMS-DT increased their total return and sharpe ratio, while their standard deviation was decreased. Furthermore, it succeeded to strengthen the profitable stability of the returns for each year, providing a real boost to the power of compounding. Since this is the

Table 8

Evaluation metrics of EWP-DT.

Metrics of EWP-DT					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
188.60	478.43	26.94	25.06	21.07	1.16

Table 9
Evaluation metrics of EWP2.

Metrics of EWP2					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
141.61	327.53	20.23	18.47	21.77	0.83

Table 10
Allocation of the five currencies in KCP-DT.

Year period	Kelly Criterion Portfolio based on AIRMS-DT (KCP-DT) (%)				
	GBP/USD	USD/JPY	EUR/USD	GBP/JPY	EUR/JPY
2010	20.00	20.00	20.00	20.00	20.00
2011	10.28	0.00	41.28	27.72	20.73
2012	4.72	59.55	8.10	14.50	13.13
2013	26.26	38.17	0.00	20.67	14.89
2014	0.00	73.27	11.54	3.52	11.67
2015	0.00	40.12	22.67	19.34	17.87
2016	47.45	24.98	27.46	0.00	0.10

Table 11
Kelly Criterion Portfolio based on AIRMS-DT (KCP-DT).

Years	2010	2011	2012	2013	2014	2015	2016
KCP-DT Returns (%)	41.60	32.73	43.20	8.03	51.64	(4.21)	47.14

Table 12
Kelly Criterion Portfolio without any AIRMS (KCP2).

Years	2010	2011	2012	2013	2014	2015	2016
KCP2 Returns (%)	33.02	18.30	44.13	1.99	43.99	(4.47)	41.39

core of the risk management logic, which can help initial investment grow exponentially.

4.3. Calibration of artificial neural networks

Due to the nature of ANN, during its application, all features considered in this study are used as input parameters. This is due to the structure of the ANN that requires enough data in order to discover patterns, which can lead to more accurate predictions. Continuously fluctuations of the number of input features combining the calibration of all the other hyperparameters would make the whole process computational costly and could lead the model to overfitting or underfitting failing develop a true predictive ability to the data.

Yearly results obtained by the implementation of the ANN-based AIRMS for each currency are depicted in Table 15. Similar to the DT model, for each FX pair the results for the year 2010 were generated using the ANN model, calibrated based on the data of the previous four years (i.e. 2006–2009). So, by means of the training and testing procedures, using the data of the first three years (75%) those of the fourth one (25%), respectively, it was succeeded to identify the best value of the models' parameters which provided the highest F1-score. Using the calibrated model, the class of the signals of the year 2010 is predicted. The trades, which produced from this forecasting, led to the return of 2010 for each currency. Similar to the first ML model, based on the walk-forward

testing routine the same procedure is followed to calculate the returns of the remaining six years. More specifically, in order to predict the return of the year 2012, the period of years 2008–2011 is used as training-testing dataset, this procedure is implemented for the period of years 2007–2010 up to 2012–2015 in order for the ANN to be calibrated appropriately the ANN model, able to predict which signals of 2016 will be transformed into an actual trade and which not. The sum of the outcome of that trades will lead to the return of 2016. In order to assess the performance of the ANN-based AIRMS, the produced returns of the system need to be compared to the original ones of the MRB channel strategy. The performance of ANN assisted MRB trading strategy and the conventional one is provided in Tables 15 and 4, respectively. In the following part of the session, a comparative presentation for each currency exchange rate is provided.

In the case of GBPUSD exchange rate, for five out of seven years the returns produced by the ANN-based AIRMS are higher or equal compared to those achieved by the MRB strategy. Specifically, the biggest increase related to the original results was achieved during the year 2010. The AIRMS succeeded to improve the return for the year 2010 by about 1375%, i.e. AIRMS achieved to turn a losing year having a negative return value of -2.97% to a winning one with 37.87% return. It is observed that AIRMS-DT succeeded to reduce the standard deviation of the returns of GBPUSD between 2010 and 2016 by 8%. The total return succeeded for the AIRMS is equal to 112.05% in comparison to the original return of 51.02% corresponds to an important increase of 119.60% .

In case of USDJPY exchange rate, for six out of seven years the returns produced by the ANN-based AIRMS are also higher or equal compared to those achieved by the MRB strategy. Specifically, the most noticeable increase related to the original returns from MRB strategy was achieved during the year 2015. The AIRMS succeeded to improve the return for the year 2015 by about 246%, i.e. from the value of 9.73% – 33.71% . Another significant raise was during the year 2016, where AIRMS-ANN almost doubled the return, i.e. from the value of 40.02% – 81.44% . The total return achieved for the AIRMS is equal to 138.43% in comparison to the original return of 74.86% , bringing off an increase of 85% .

Accordingly, in the case of EURUSD exchange rate, for five out of seven years the returns produced by the ANN-based AIRMS are higher or equal compared to those achieved by the MRB strategy. Specifically, the largest increase related to the original results was achieved during the year 2013. The AIRMS succeeded to improve the return for the year 2013 significantly, i.e. AIRMS transformed an almost zero return to a strongly positive one of 26.84% . AIRMS-ANN achieved an impressive decrease in the standard deviation of the return of EURUSD by 42%, improving the risk profile of this currency dramatically.

In the case of GBPJPY exchange rate, for six out of seven years the returns produced by AIRMS based on ANN are higher or equal

Table 13
Evaluation metrics of KCP-DT.

Metrics of KCP-DT					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
220.14	621.46	31.44	29.82	19.70	1.47

Table 14
Evaluation metrics of KCP2.

Metrics of KCP2					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
178.35	449.89	25.48	23.97	20.42	1.14

Table 15
Currency returns based on AIRMS-ANN.

Years	Returns via AIRMS-ANN (%)				
	GBPUSD	USDJPY	EURUSD	GBPJPY	EURJPY
2010	37.87	(50.03)	66.40	41.52	43.23
2011	(19.20)	(22.72)	0.08	20.06	79.70
2012	5.95	39.96	25.24	54.44	59.62
2013	(16.24)	13.25	26.84	10.83	(32.98)
2014	(4.63)	42.82	36.35	69.10	34.60
2015	11.01	33.71	15.23	(13.49)	(21.37)
2016	97.30	81.44	(17.70)	42.40	25.78
Total return (%)	112.05	138.43	152.44	224.85	188.57

compared to those achieved by the MRB strategy. Specifically, the highest increase related to the original results was observed during the year 2015. The AIRMS succeeded to improve the return for the year 2015 by about 53.70%, i.e. from −29.14 to −13.49%. Related to the standard deviation of the returns of GBPJPY, AIRMS-ANN achieved a significant reduction by 22.7%. Also, the total return succeeded for the AIRMS is equal to 224.85% in comparison to the original return of 217.59% corresponds to an increase of 3%.

In the case of EURJPY exchange rate, for three out of seven the returns generated by the ANN-based AIRMS are higher or equal to the original strategy. Specifically, the biggest increase related to the original results was observed during the year 2015. The AIRMS succeeded to improve the return for the year 2015 by about 57.22%, reducing to the half a negative year from −49.95 to −21.70%. It can be seen, AIRMS-ANN provided more stable returns through the period of the 7 years reducing the standard deviation of the returns by 16.6%.

Concluding this part of the investigation, worth mentioning that ANN based AIRMS succeeded to ameliorate the risk profile of the majority of the currency pairs. It reduced the standard deviation of four out of the five pairs and increased the total return the three of them. These observations prove that ANN based AIRMS has a satisfactory ability to predict the class of the produced signals in such a way in order to generate profits. The most significant fact is that the developed risk management tool based on ANN achieved to improve almost all of the negative returns that were produced by the MRB channel strategy. The reduction of the negative performances of the trading strategy will be an important advantage at the portfolio construction stage.

4.4. Re-Construction of EWP2 and KCP2 based on AIRMS-ANN

In the previous section, the returns for each currency pair, which were produced by ANN based AIRMS, were presented. In this section, the two portfolios constructed in previous study by Chandrinós and Lagaros (2018) are reconstructed. The first one is a simple average portfolio where all pairs have the same weight coefficient. The second portfolio is rather more advanced while different weight coefficients are assigned to the various currency pairs. The weights of each pair for each year are determined from its performance during the previous year based on Kelly Criterion.

Table 16
Allocation of the five currencies in EWP-ANN.

Year period	Equally Weighted Portfolio based on AIRMS-ANN (EWP-ANN) (%)				
	GBP/USD	USD/JPY	EUR/USD	GBP/JPY	EUR/JPY
2010	20.00	20.00	20.00	20.00	20.00
2011	20.00	20.00	20.00	20.00	20.00
2012	20.00	20.00	20.00	20.00	20.00
2013	20.00	20.00	20.00	20.00	20.00
2014	20.00	20.00	20.00	20.00	20.00
2015	20.00	20.00	20.00	20.00	20.00
2016	20.00	20.00	20.00	20.00	20.00

Table 17
Equally Weighted Portfolio based on AIRMS-ANN (EWP-ANN).

Years	2010	2011	2012	2013	2014	2015	2016
EWP-ANN Returns (%)	27.80	11.58	37.04	0.34	35.65	5.02	45.84

The portfolios are formed based on the exact method that was used for EWP2 and KCP2 by Chandrinós and Lagaros (2018). These two re-constructed portfolios will be denoted as EWP-ANN and KCP-ANN accordingly. In order to evaluate the effect of the AIRMS based on ANN on the performance of the portfolios, the EWP-ANN and KCP-ANN will be compared to the original EWP2 and KCP2.

Table 16 shows the yearly proportions of each currency pair assigned to the EWP-ANN. The allocation chosen is equal to 1/5 since the EWP-ANN is composed of five currency pairs. Table 17 contains the returns of the EWP-ANN, which is based on the AIRMS-ANN, while Table 7 presents the returns of the EWP2 from the original MRB channel strategy. Comparing the results of these two tables it can be observed that the re-constructed equally weighted portfolio succeeds to outperform the EWP2 for five out of seven years. Specifically, the broadest increase related to the original results was achieved during the year 2015. EWP-ANN portfolio succeeded to improve the return of 2015 by about 160%, i.e. the negative return of −8.32% was transformed into a profitable one of 5.02%.

These improvements that the developed ANN based risk management system achieved on the performance of EWP2 are more presented when comparing the evaluation metrics of the two equally weighted portfolios. These metrics can be seen in Tables 18 and 9, which contain the total return, the compound growth, the arithmetic average, geometric return, standard deviation and the Sharpe ratio of each portfolio corresponding. Notably, the EWP-ANN portfolio attained a total return of 163.27% improving the total return of the original equally weighted portfolio by approximate 15.30%. Regarding compound growth, the re-constructed equally weighted portfolio reported a return equal to 407% that compared the EWP2's compound growth is increased by 24.40%. This means that an investor who invests \$10,000 in 2010 at EWP-DT, he would have \$40,700 by the end of 2016. Most importantly, the AIRMS-ANN improved the two most significant metrics related to risk management field of the new portfolio; the standard deviation and the sharpe ratio. The standard deviation of the new average port-

Table 18
Evaluation metrics of EWP-ANN.

Metrics of EWP-ANN					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
163.27	407.38	23.32	22.22	16.34	1.27

Table 19
Allocation of the five currencies in KCP-ANN.

Year period	Kelly Criterion Portfolio based on AIRMS-ANN (KCP-ANN) (%)				
	GBP/USD	USD/JPY	EUR/USD	GBP/JPY	EUR/JPY
2010	20.00	20.00	20.00	20.00	20.00
2011	16.37	0.00	39.68	32.29	11.66
2012	0.10	0.00	0.10	52.10	47.70
2013	4.28	19.21	17.69	37.18	21.64
2014	0.00	27.98	70.46	1.56	0.00
2015	0.00	41.52	25.12	20.60	12.76
2016	28.97	34.37	36.66	0.00	0.00

Table 20
Kelly Criterion Portfolio based on AIRMS-ANN (KCP-ANN).

Years	2010	2011	2012	2013	2014	2015	2016
KCP-ANN Returns (%)	27.80	11.58	37.04	0.34	35.65	5.02	45.84

folio decreased significantly by 25%, from 21.77% to 16.34%, making the EWP-ANN less risky. Finally, the application of the AIRMS-ANN to the EWP2 resulted in an impressive increase of the sharpe ratio by 56.50%, i.e. from 0.81 to 1.27.

Table 19 presents the yearly percentages of each currency pair assigned to the KCP-ANN. This allocation during a year is based on the Kelly Criterion of each pair from the previous year. For this reason, the 2010's weight coefficients are equal for all the currencies due to the lack of data from 2009. Table 20 contains the returns of the KCP-ANN portfolio, which is based on the AIRMS-ANN, while Table 12 presents the returns of the KCP2 from the original MRB channel strategy. Specifically, the largest increase related to the original returns was during 2015. The KCP-ANN succeeded to raise the return of 2015 by about 375%, i.e. AIRMS succeeded to turn a losing year with -4.47 to 12.31% . Another significant amelioration which the KC-based portfolio succeeded by AIRMS-ANN was during 2016, where increased even more an already high return of 41.39 – 49.69% . This example proves the predictions of the AIRMS-ANN are statistically significant because is even more difficult for a system to improve an already good result than to increase a negative outcome.

These improvements that the developed ANN based risk management system achieved on the performance of KCP2 are more clearly presented when comparing the evaluation metrics of the two KC based portfolios. These metrics can be seen in Tables 21 and 14, which contain evaluation metrics of each portfolio.

Table 21
Evaluation metrics of KCP-ANN.

Metrics of KCP-ANN					
Total return (%)	Compound growth (%)	Arithmetic average (%)	Geometric return (%)	Standard deviation (%)	Sharpe ratio
201.47	544.84	28.78	27.40	18.88	1.39

lio. Notably, the KCP-ANN portfolio achieved significantly the risk evaluation metrics; standard deviation and sharpe ratio. The standard deviation of the new KC portfolio was lower by 7.50%, from 20.42 to 18.89% and the sharpe ratio increased by 23.70%, i.e. from 1.14 to 1.39. Furthermore, the high total return of 201.47% improving the total return of the original KC based portfolio by approximately 13.00%. Regarding compound growth, the updated portfolio based on KC reported a return equal to 544.84% that compared to the EWP2's compound growth is increased by 21%. This means that an investor who invests \$10,000 in 2010 at EWP-ANN, he would have \$54,480 by the end of 2016.

The contentious improvements on the performance of the EWP2 and KCP2 portfolios, that the adjustment of the AIRMS-ANN offered, underline that its ability to recognize the quality of the signal produced is significant. In both portfolios, the use of AIRMS-ANN increased their total return and sharpe ratio, while their standard deviation was decreased. It succeeded to strengthen the profitable stability of the returns each year giving a real boost to the power of compounding by creating two portfolios with seven consequent profitable years. Warren Buffett, responding to an interview question for the single most powerful factor behind his investing success answered "compound interest". According to this factor, this great investor had developed an investment advice, a motto: "The first rule is not to lose money. The second rule is not to forget the first rule."

4.5. Comparison of the performance of the application of the ANN and DT

Firstly, the effect of the utilization of AIRMS-DT and AIRMS-ANN are compared in the performance of the five currencies. For GBP/USD, as it can be seen in Fig. 4, both AIRMS succeeded to outperform the total return of the MRB strategy. A notable fact is that the AIRMS-ANN achieved to maintain always positive total return after any given trade during the entire seven-year period. For USD/JPY, as it can be observed in Fig. 5, both AIRMS produced higher profits than the original strategy. Furthermore, the AIRMS-DT succeeded to decrease significantly the maximum drawdown during 2010–2016. For EUR/USD, in Fig. 6, the AIRMS-DT achieved to raise the total return of the initial trading strategy, while the AIRMS-ANN performed similarly to the MRB strategy. For GBP/JPY, in Fig. 7, the AIRMS increased the total return, while the AIRMS-ANN reduced significantly the drawdowns during this seven-year period. For EUR/JPY, as it can be seen Fig. 8, the AIRMS-DT not only rose the total return but also achieve to reduce the average drawdown of the MRB strategy, while the AIRMS-ANN underperformed slightly the original return.

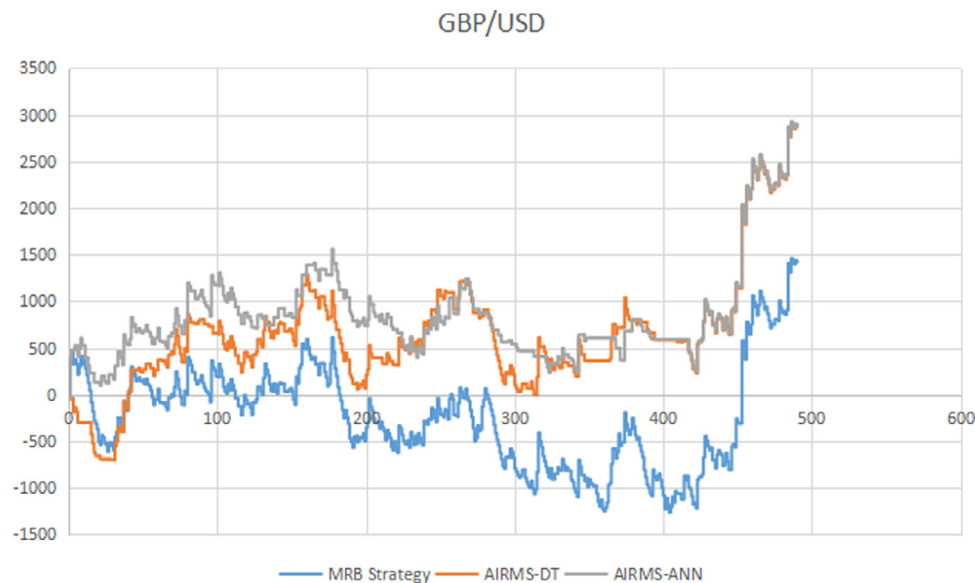


Fig. 4. Representation of the performance per trade of GBP/USD between 2010 and 2016 based on MRB strategy, AIRMS-DT and AIRMS-ANN.

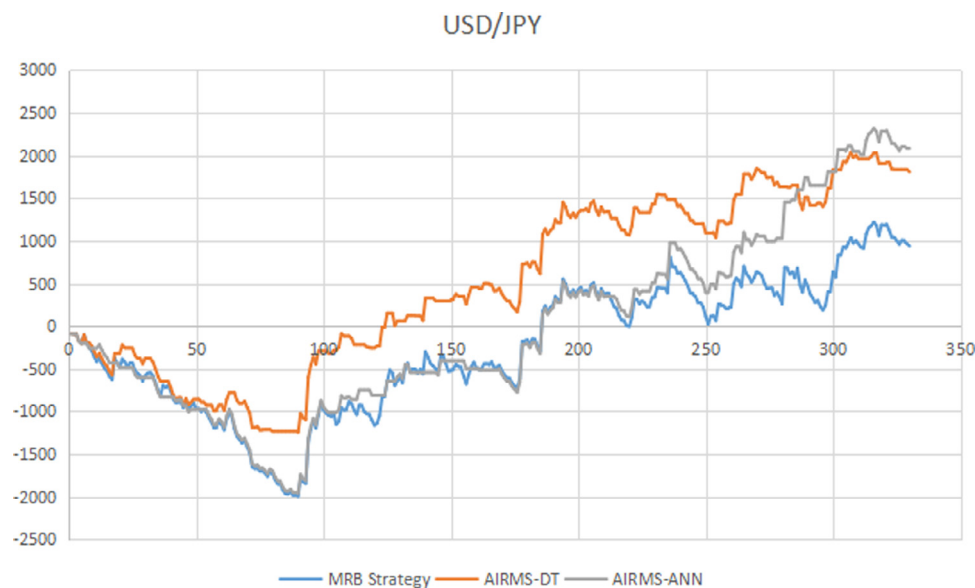


Fig. 5. Representation of the performance per trade of USD/JPY between 2010 and 2016 based on MRB strategy, AIRMS-DT and AIRMS-ANN.

Then, the performance of the equally weighted and Kelly criterion portfolios based on AIRMS-ANN and AIRMS-DT are compared. For EWPs, as it can be observed in Fig. 9, the EWP-DT and EWP-ANN beat the original EWP2, while EWP based on DT outperformed the one based on ANN in terms of total return. However, the AIRMS-ANN succeeded to construct an EWP with lower standard deviation and higher sharpe ratio than EWP-DT. For KCPs, in Fig. 10, the KCP-DT and KCP-ANN achieved higher total return during the seven-year period than the original KCP2, while the KCP-DT beat the KCP-ANN related to total return. Nevertheless, the use of AIRMS-ANN led to a KCP with a lower standard deviation than the KCP-DT.

4.6. Technical observations and applicability

The experimental stage of the study resulted in several noteworthy technical observations. It is known from the literature that data noise is a major constraint for applying ANNs to financial data and also that ANNs depict poor performance due to the absence of a specific structure of financial data. In the current study, aiming to deal with these issues, ANNs were applied to financial data that were already modified under the special trading tool named Modified Renko Bars (MRBs), developed in previous study (Chandrinis & Lagaros, 2018). This is not a common smoothing method such as moving averages etc. MRBs transforms time-based charts into ones that are based only on price movement. The use of MRBs reduces dramatically the noise of financial data, making easier for the ML algorithms to process inputs. This observation was very critical and

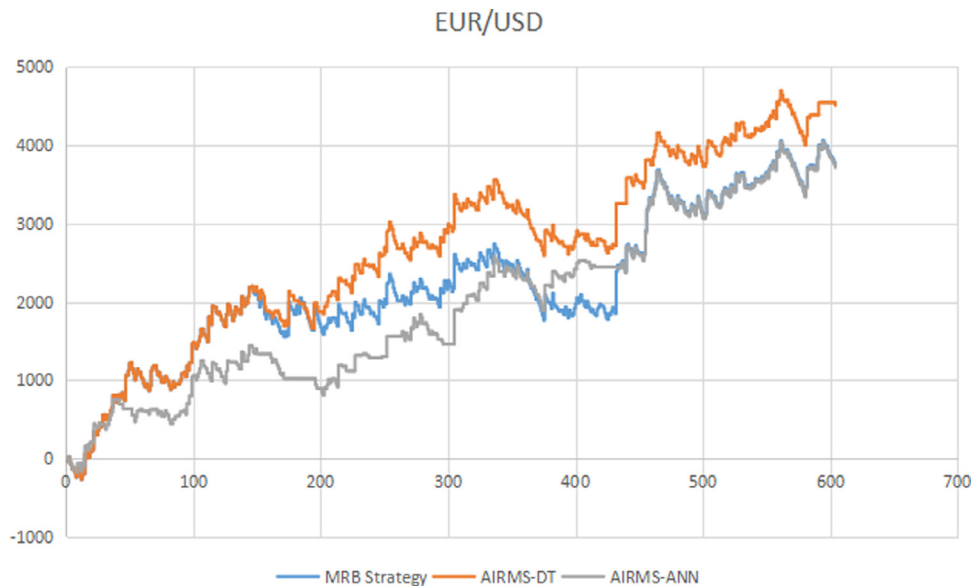


Fig. 6. Representation of the performance per trade of EUR/USD between 2010 and 2016 based on MRB strategy, AIRMS-DT and AIRMS-ANN.

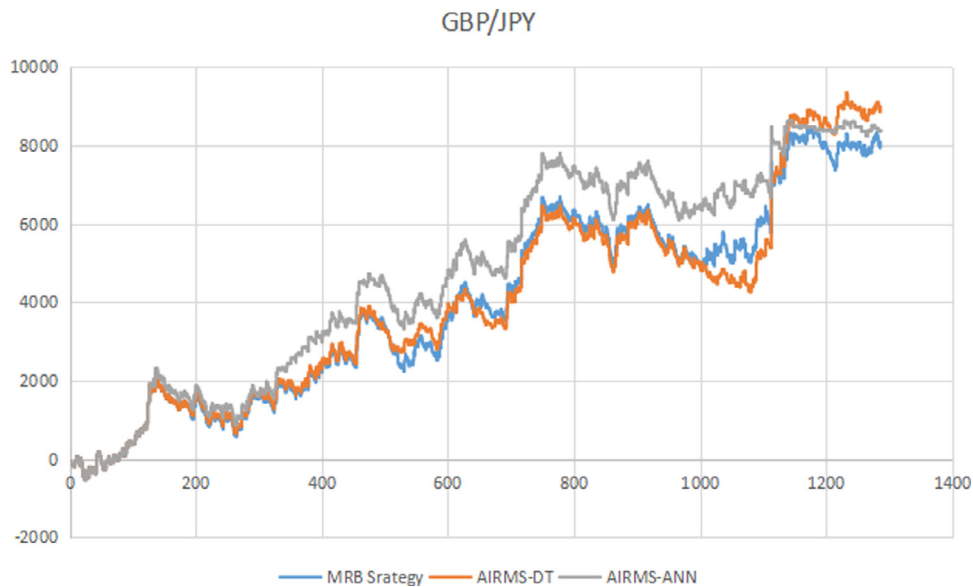


Fig. 7. Representation of the performance per trade of GBP/JPY between 2010 and 2016 based on MRB strategy, AIRMS-DT and AIRMS-ANN.

helpful for ANNs to produce the significant results described previously. Worth mentioning also that MRBs are applicable to any financial data and so it can become a valuable component for improving the ability of ANNs to comprehend the income data from financial assets.

Furthermore, the evaluation metric that denotes how well each model fits the trained data were adjusted. Predicting the trend of a financial asset or the quality of produced signals from a strategy is not a simple binary decision: only up and down or profitable and no profitable, respectively. There is another dimension: “how strong the trend or the signal is?”; hence, the responses are not equally weighted. Therefore, instead of using accuracy between predictions and targets as a metric, F1-score is applied; which, as explained in the previous section, is able to capture the nature of

this type of responses. This is another tool that it can be used to support and ameliorate further the analysis of financial data by expert systems.

Finally, during the calibration procedure of ANNs, it is observed that the optimization field of their parameters are very wide: the number of the hidden layers, the number of the neuron of each layer, the direction of the information (feedforward or recurrent neural networks), the number and the nature of the input, the amount of the observations. This fact is the main limitation of utilizing ANNs in this work, since it may be a hindering factor of the system’s swift deployment in real-world applications. In the future, we target to investigate the ability to optimize the components of the neural networks in depth.

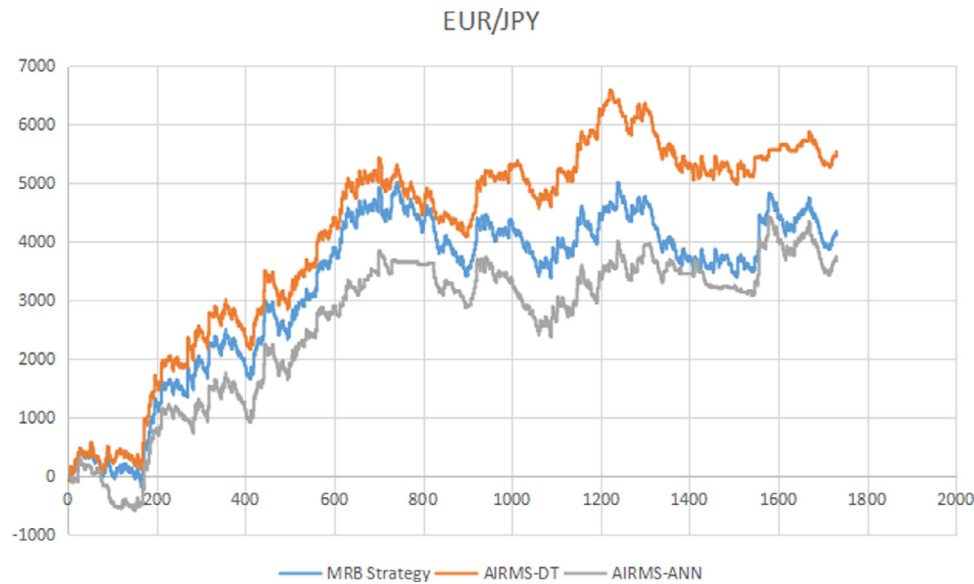


Fig. 8. Representation of the performance per trade of EUR/JPY between 2010 and 2016 based on MRB strategy, AIRMS-DT and AIRMS-ANN.

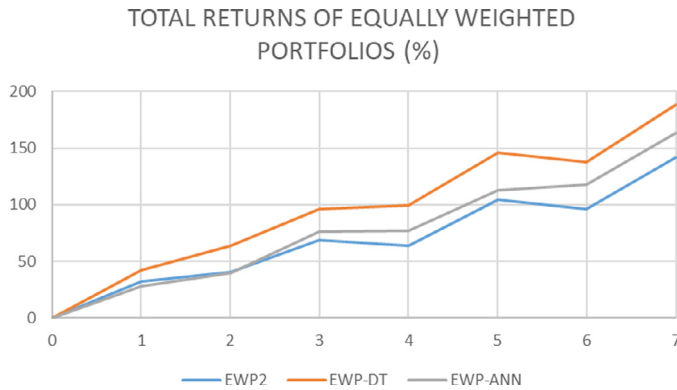


Fig. 9. Representation of the performance of the equally weighted portfolios based on MRB strategy, AIRMS-DT and AIRMS-ANN.

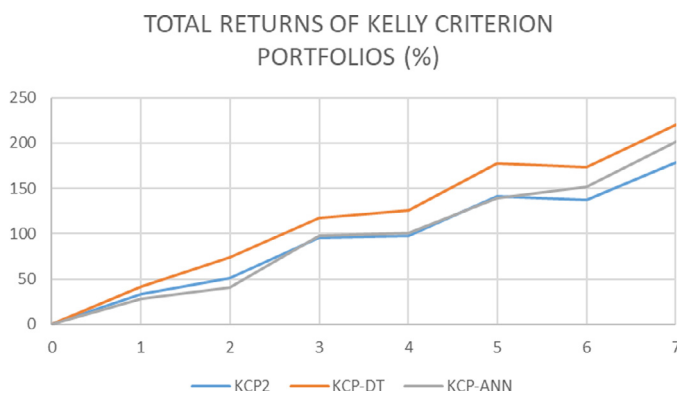


Fig. 10. Representation of the performance of the Kelly Criterion portfolios based on MRB strategy, AIRMS-DT and AIRMS-ANN.

5. Conclusion

The major achievement of this study was the successful application of two machine learning models for risk management purposes; resulting in the development of an artificial intelligent risk management system (AIRMS). Decision trees (DT) and artificial neural networks (ANN) are the two models that were used in

this work. In previous work (Chandrinos & Lagaros, 2018), the authors proposed two profitable currency portfolios (labeled as EWP2 and KCP2) that were based on the optimized channel MRB strategy. These two portfolios are constituted by five currencies exchange rates (GBP/USD, USD/JPY, EUR/USD, GBP/JPY and EUR/JPY) and their returns refer to a seven-year period, from the year 2010 to 2016.

The main target of the developed AIRMS was to improve the performance of the two portfolios not by generating new profits but preventing them from losses. Therefore, AIRMS was applied to the optimized trading strategy in order to recognize which signals will be profitable and which ones will not, transforming the profitable ones into trades. For both risk management systems developed in the current study (i.e. DT and ANN based ones, labeled as AIRMS-DT and AIRMS-ANN, respectively) the dynamic moving window was used in order to train, test and finally predict the quality of the produced signals of the next year. Worth mentioning that the way that the two machine learning models were calibrated is crucial for the results produced and that the two systems succeeded to provide significant amelioration on the performance of both EWP2 and KCP2 currency portfolios, resulting into improved portfolios that can be even more attractive to potential investors. The two systems not only increased the profitability of both portfolios but also mainly managed to further strengthen their sharpe ratio by reducing the standard deviation.

Analysing the results that AIRMS-DT and AIRMS-ANN provided, it was observed the stability of their predicting performance that is a crucial importance characteristic for machine learning models. By its definition, the ultimate goal of the two risk management systems was difficult; since they were employed in order to increase the performance of two already significantly profitable portfolios and they succeeded both in terms of total return and sharpe ratio. DT based AIRMS firstly managed to launch the total return of each of the currencies increasing it by an average rate of 50%. This fact had as a result of the developed DT system to improve even more the high total return and the sharpe ratio decreasing at the same time the standard deviation for both portfolios. Similarly, ANN-based AIRMS managed to raise the total return of the five currencies by an averagely 40% succeeding the same impressive results to during the constructions of both portfolios.

A noteworthy evidence of the successful application of the DTs and ANNs machine learning techniques as risk management tools was their ability not only to slightly improve negative returns but

to turn most of the losing years into profitable ones and also to increase even more already highly profitable years. All these facts proved undeniably the significance of the results of this study. They proved that this study achieved to utilize successfully these two machine learning algorithms for a quite innovative objective; to control the risk of investment portfolios. Another notable observation is that comparing the two AIRMS, the utilization of DT constantly led to higher returns, while the ANN succeeded in both types of the constructed portfolios lower standard deviations.

Contrary to previous studies that employed ML algorithms in financial data, where mainly focused on predicting the actual price of stocks and currency pairs or the general direction of them, in the current study a completely different approach is introduced. It is proved that ML can be utilized efficiently for risk management of currency portfolio strategies. This study introduces a new risk management tool to the expert systems community that can be applied to any investment/trading strategy of any financial instrument improving its expected return (predicting on average the quality of future trades) and hence the profitability of the investor.

Future work should definitely include working on more diverse strategies since the advantage of the proposed system is that it does not adhere to a specific strategy. Only data from the strategies' former results are needed to train the models. This characteristic makes AIRMS very volatile and therefore, it could be used in a variety of strategies (such as technical strategies based on moving averages or relative strength index, momentum strategies relying on past returns). Furthermore, feature extraction could be more thoroughly examined; although an area of more than 20 features was covered, some completely different ones could be a game changer. In addition, AIRMs will be applied to stocks (to the total number of stocks of an index such as S&P500). This will provide another statistically significant test for the prediction ability of the developed intelligent risk management systems. Also, it will give us the chance for a direct comparison to relative work by other researchers that mostly focus on stocks. Finally, different ML algo-

gorithms such SVMs and more sophisticated architectures of ANNs (such as Recurrent Neural Networks), that have presented significant results in other research fields, e.g. computer vision and natural language processing, will be examined. Particularly, SVMs could be an interesting path to follow, since they require a lot less fine-tuning and training times than ANNs.

References

- Arevalo, R., Garcia, J., Guijarro, F., & Peris, A. (2017). A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. *Expert Systems with Applications*, 81, 177–192.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. California: Belmont.
- Chandrinos, S. K., & Lagaros, N. D. (2018). Construction of currency portfolios by means of an optimized investment strategy. *Operations Research Perspectives*, 5, 32–44.
- Chang, P., Liu, C., Lin, J., Fan, C., & Ng, C. (2009). A neural network with a case based dynamic window for stock trading prediction. *Expert Systems with Applications*, 36, 6889–6898.
- Chen, A., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial markets: Forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30, 901–923.
- Choudhry, R., & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. In *Proceedings of the world academy of science, engineering and technology* 29.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multi-layer feedforward networks are universal approximators. *Neural Networks*, 2, 359–366.
- Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990). Stock market prediction system with modular neural networks. In *Proceedings of the IEEE international joint conference on neural networks*.
- Li, Y., Fu, Y., Li, H., & Zhang, S.-W. (2009). The improved training algorithm of back propagation neural network with self-adaptive learning rate. In *Proceedings of the international conference on computational intelligence and natural computing* (pp. 73–76). 1.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7, 77–91.
- Refenes, A. N., Zaprani, A., & Francis, G. (1994). Stock performance modeling using neural networks: A comparative study with regression models. *Neural Networks*, 7(2), 375–388.
- Teixeira, L. A., & Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems and Applications*, 37, 6885–6890.
- Wu, M., Lin, S., & Lin, C. (2006). An effective application of decision tree to stock trading. *Expert Systems with Applications*, 31, 270–274.