科技部專題研究計畫申請書

申請條碼：109WFA0910553

一、基本資料：

計畫類別（單選）	一般研究計畫				
研究型別	個別型				
計畫歸屬	人文司				
申請機構/系所（單位）	國立成功大學工業與資訊管理學系（所）				
本計畫主持人姓名	李昇暾	職稱	教授	身分證號碼	R12012****
本計畫名稱 中文	單調性極限學習機模型於單類別分類之研究				
本計畫名稱 英文	A Novel Monotonic Extreme Learning Machine for One-class Classification				
整合型總計畫名稱					
整合型總計畫主持人				身分證號碼	
全程執行期限	自民國 109 年 08 月 01 日起至民國 110 年 07 月 31 日				
研究學門	學門代碼		學門名稱		
	H54A1		資訊管理		

【請考量己身負荷，申請適量計畫】
本年度申請主持科技部各類研究計畫(含預核案)共 1 件。(共同主持之計畫不予計入)

本計畫是否為國際合作研究：■否； □是，請加填表IM01~IM03

本計畫是否申請海洋研究船：■否； □是，請務必填寫表CM15。

1. 本計畫是否有進行下列實驗/研究：（勾選下列任一項，須附相關實驗/研究同意文件）
□人體試驗/人體檢體 □人類胚胎/人類胚胎幹細胞 □基因重組實驗 □基因轉殖田間試驗 □第二級以上感染性生物材料 □動物實驗(須同時加附動物實驗倫理3R說明)
2. 本計畫是否為人文司行為科學研究計畫：□是(請檢附已送研究倫理審查之證明文件)； ■否
3. 本計畫是否為臨床試驗研究計畫：□是(請增填性別分析檢核表CM16)； ■否

計畫連絡人	姓名： 李昇暾 電話：(公) 06-2757575-53126 (宅/手機) 06-2757575-53126		
通訊地址	台南市大學路1號成功大學工業與資訊管理學系		
傳真號碼	06-2362162	E-MAIL	stli@mail.ncku.edu.tw

計畫主持人簽章：______________　　日期：______________

二、研究計畫中英文摘要：請就本計畫要點作一概述，並依本計畫性質自訂關鍵詞。

計畫中文關鍵詞	單類別分類、極限學習機、單調性分類法、資料探勘、機器學習
計畫英文關鍵詞	One-class Classification, Extreme Learning Machine, Monotonicity Classification, Data Mining, Machine Learning
計畫中文摘要	分類問題在機器學習領域中是一重要議題，因其可以透過數據特徵與相關性，進一步輔助決策者進行更精確之管理與預測。分類問題主要目標在於利用一些樣本資料訓練分類器，使訓練好的分類器能針對未見過的資料進行分類預測。在眾多的分類模型之中，極限學習機(ELM)是一種很特別的分類法。ELM透過隨機指派輸入層的權重與偏誤值，使其擁有比傳統演算法快上數倍的訓練時間，並擁有較好的泛化能力。另一方面，若將先驗知識如單調性限制加入分類模型與決策中，則能提高分類精確度，並減少訓練時間。然而，對ELM之單層前饋式神經網路結構而言，並不具備特徵學習與抽象的能力，因此，處理具先驗知識分類問題的效果並不好，故文獻上少有探討單調性限制結合ELM之研究。 過往研究曾有將單調性限制式套用到原始ELM，並將數學模型轉換為二次規畫問題，其中考慮了誤差和單調性。然而，修正後的模型已進一步變更ELM內部模型之結構。因此，有必要針對此部分缺失進行改善。另外，傳統的多類別分類中，若新數據不符合任一類別，將會造成錯誤分類並影響準確度。基於此，單類別分類法區分數據中的目標類及異常值類，其僅使用單類別標籤資料進行訓練，成為近年來熱門的研究主題之一，並已廣泛地應用於異常值偵測與機台故障檢測等應用上。 有鑑於此，本研究目標希望能建立一個針對單調性資料進行單類別分類的分類模型。因此，我們提出了一個基於ELM的單調性單類別分類器，透過內部權重的調整來滿足單調性的限制式並且維持原始ELM結構，最後則建立一個決策函數來進行單類別的判斷。預期所提的模型將可顯著提高分類的準確性與效率。
計畫英文摘要	The classification problem is an important task in the field of machine learning since it can help make predictions through the characteristics and relevance of data. Its goal is to use some sample data to train the classifier, so that the classifier can classify unknown data into one of several pre-defined categories. Among them, Extreme Learning Machine (ELM) is a classification model that is often used. By the characteristic of randomly determining the weight, ELM can greatly reduce the computation time, making it have several times faster training time than traditional algorithms, and has better generalization capabilities. Besides, adding prior knowledge, such as monotonic constraints, to classification models and decision-making can improve classification accuracy and reduce training time. However, for the Single Layer Feed-forward Neural Network (SLFN) structure of ELM, it does not have the capability of feature learning and abstraction. Therefore, ELM is not a good model for monotonic classification problems, so there have been few studies that work on monotonic classification based on ELM network structure. Previous studies have applied the monotonic constraint to the original ELM. It satisfied the monotonicity constraints by imposing the constraints to the original ELM, however, the structure of the original ELM model then had been changed. Thus, it is necessary to improve the lack of this part. On the other hand, in traditional multi-class classification, if the unknown data does not meet any one category, it will cause misclassification and affect accuracy. Different from them, one-class classification tries to distinguish between objects from target class and outlier class. One-class classification is widely used in outlier detection and machine fault detection. In view of this, the goal of this project is to establish a classification model for one-class classification of monotonic data. Therefore, we propose a monotonic

	one-class classifier based on ELM. We satisfy the monotonicity constraints by altering the expression of the network weights and generate a decision function to handle the problem of one-class classification. It is expected that the proposed model will significantly improve the accuracy and efficiency of classification.
計畫概述	請概述執行本計畫之目的及可能產生對社會、經濟、學術發展等面向的預期影響性(三百字以內)。 ※此部分內容於獲核定補助後將逕予公開
	本計劃將建構一個針對單調性資料進行單類別分類的極限學習機分類模型，不僅將相關領域知識納入考慮，也根據現今興起的大數據和機器學習議題，進行演算法設計與推導，將之發展成適用於單調性資料進行單類別分類的極限學習機分類模型。因此，本模型預期透過內部權重的調整來滿足單調性的限制式並且維持原始ELM結構，並建立一個決策函數來進行單類別的判斷，將可運用至實務案例中。此項研究成果亦將有助於提升機器學習與大數據分析領域之學術成果。

三、研究計畫內容（以中文或英文撰寫）：

（一）研究計畫之背景。請詳述本研究計畫所要探討或解決的問題、研究原創性、重要性、預期影響性及國內外有關本計畫之研究情況、重要參考文獻之評述等。如為連續性計畫應說明上年度研究進度。

（二）研究方法、進行步驟及執行進度。請分年列述：1.本計畫採用之研究方法與原因及其創新性。2.預計可能遭遇之困難及解決途徑。3.重要儀器之配合使用情形。4.如為須赴國外或大陸地區研究，請詳述其必要性以及預期效益等。

（三）預期完成之工作項目及成果。請分年列述：1.預期完成之工作項目。2.對於參與之工作人員，預期可獲之訓練。3.預期完成之研究成果（如實務應用績效、期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4.學術研究、國家發展及其他應用方面預期之貢獻。

（四）整合型研究計畫說明。如為整合型研究計畫請就以上各點分別說明與其他子計畫之相關性。

1. Introduction

1.1 Research Background

Training a classifier by several samples, and predicts the class label of an unseen sample by the trained classifier is a fundamental task of classification problem. Extreme learning machine (ELM) is a popular classifier when encountering a classification problem (Huang, Zhu, & Siew, 2006). The ELM algorithm randomly assigned the input weights and hidden layer biases so that it can learn thousands of times faster than the BP-based algorithm in conventional neural networks. In addition, ELM can produce good generalization performance in most cases (Zhu, Tsang, Wang, & Ashfaq, 2017).

One-class classification (OCC) (Khan & Madden, 2009) has received much interest during recent years. Different from normal classification, training data from only one class (target class), are well characterized or labeled, while there are no or few samples from the other class (outlier class). OCC has been widely applied in several fields such as anomaly detection (Perdisci, Gu, & Lee, 2006) (Li, Huang, Tian, & Xu, 2003) and machine failure detection (Shin, Eom, & Kim, 2005). Furthermore, using the one class data as training data for multi-class classifier would generate a large number of false positive detection and reduce the accuracy instead (Tax, 2002).

Classification problems where there is background knowledge in the form of ordinal evaluations are very common. A particular case of ordinal classification is monotonic classification (Gutiérrez & García, 2016). Monotonicity constraints require that the class label assigned to a pattern should be greater or equal to the class labels assigned to the patterns it dominates. Monotonic classification problems widely exist in a real-world situation such as disease diagnoses, medical care and the relationship between distance and time.

1.2 Research Motivation

It is necessary to add monotonicity constraints to the study. By adding monotonicity constraints, the accuracy of the classification problem with the input data is known to be monotonic can be improved. So far, there have been few studies that work on monotonic classification based on ELM network structure. Zhu et al. (2017) proposed a monotonic classification ELM, however, they satisfied the monotonicity constraints by imposing the constraints to the original ELM. The structure of the original ELM model then had been changed. Different from them, we satisfy the monotonicity by adjusting the weights into exponential forms, so we can maintain the original network structure of ELM and achieve the goal of monotonic classification.

Simultaneously, due to the necessity of classification research for one-class data, we refer to the method proposed by Leng, Qi, Miao, Zhu, and Su (2015) to establish a distance-related decision function to classify one-class data.

Overall, this project proposes a monotonic one-class classifier model based on ELM. We satisfy the monotonic constraint by altering the weight of the neural network and classify the one-class data by using a decision function; the detail of the method will be described in subsequent sections.

1.3 Research Objectives

The main idea of this project is to construct a one-class classification model to predict monotonic data. More precisely, the dataset that has been confirmed to be monotonic is used as the input to construct the classifier. ELM is adopted as the neural network structure in this research. We modify the weights of the network to satisfy the monotonicity constraints and build up a decision function to perform one-class classification.

The purpose of this project is to propose an innovative ELM model that satisfies monotonicity constraints in structure. We expect this approach can significantly improve the accuracy and efficiency of traditional ELM processing the one-class monotonic data classification.

2. Literature Review

2.1 Extreme Learning Machine (ELM)

2.1.1 Introduction of ELM

Extreme Learning Machine (ELM) is introduced by Huang, Zhu, and Siew (2004). ELM is a classification and regression model that is based on a Single Layer Feed-forward Neural Network (SLFN) with random weights and bias values. The random assigned weights and bias values are also the main idea of ELM. The traditional learning algorithm in the neural network such as backpropagation requires setting several parameters and may fall into the local minimum sometimes. However, ELM only requires setting the number of neurons in the hidden layer and the activation function. It doesn’t have to adjust or revise the input weights and the bias, and it also comes up with only one solution. After the training process in ELM, since weights and deviations have been determined, the network output parameters can be quickly and automatically calculated using an optimization scheme involving the target vectors. Therefore, ELM has a faster learning speed and is relatively advantageous in good generalization performance. The ELM’s network structure is shown in Figure 1 (Tissera & McDonnell, 2016).

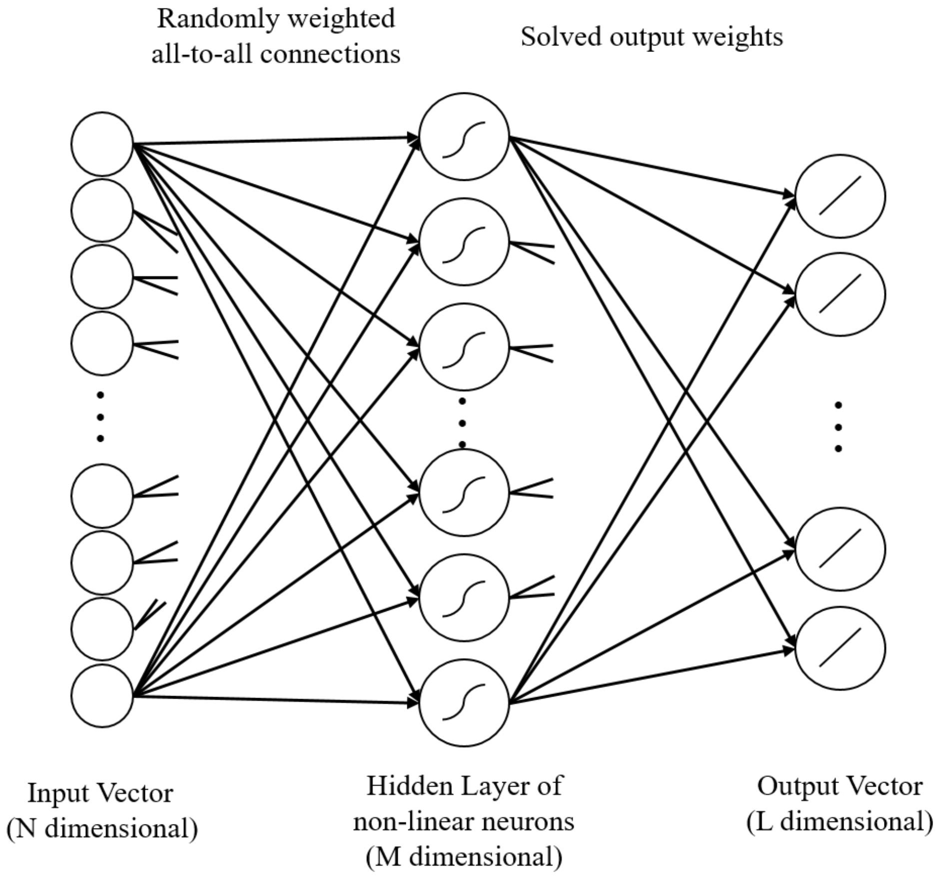

Figure 1. ELM Model

2.1.2 Mathematical of ELM

The performance of the network tends to be better if the norm of the weights are small (Bartlett, 1998). ELM aims to reach not only the smallest training error but also the smallest norm of output weights (Huang, Zhu, et al., 2006) between the hidden layer and the output layer. Thus, better generalization performance can be expected for ELM networks.

Given N training data $(\mathbf{x}_i, t_i)_{i=1}^{N}$, where $\mathbf{x}_i = [x_{i1}, \cdots, x_{in}]^T \in R^n$ is the individual feature vector with dimension n and $t_i \in R^m$ is the desired target output. The ELM output function can be formulated as

$$f(x) = \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} = \sum_{j=1}^{L} \beta_j G(\mathbf{w}_j, b_j, \mathbf{x}) \tag{1}$$

where, $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_L]^T$ is the vector of the output weights between the hidden layer and the output layer, $\mathbf{w_j} = [w_{j1}, w_{j2}, ..., w_{jn}]^T$ denotes the input weights between input nodes and hidden nodes, b_j is the bias of j th hidden node, $\mathbf{h(x)} = [G(\mathbf{w}_1, b_1, \mathbf{x}), ..., G(\mathbf{w}_L, b_L, \mathbf{x})]^T$ is the output vector of the hidden layer with respect to input $\mathbf{x}$, and $G(\mathbf{w}, b, \mathbf{x})$ is the activation function. The goal of the ELM classification problem can be formulated as

$$\begin{aligned} &\min\ L_{P_{ELM}} = \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \\ &\text{s.t.}\quad h(\mathrm{x_i})^T\beta = t_i - \xi_i \end{aligned} \tag{2}$$

where ξ_i is the slack variable of the training sample $\mathbf{x}_i$ and C controls the tradeoff between the output weights and the errors. Based on the Karush-Kuhn-Tucker (KKT) theorem (Huang & Chen, 2007), training ELM is equivalent to solve the following dual optimization problem:

$$L_{D_{ELM}} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|\xi_i\|^2 - \sum_{i=1}^{N}\alpha_i(\mathbf{h}(\mathbf{x}_i)^T\boldsymbol{\beta} - t_i + \xi_i) \tag{3}$$

We can have the KKT optimality conditions of (3) as follows:

$$\frac{\partial L_{D_{ELM}}}{\partial \boldsymbol{\beta}} = 0 \to \boldsymbol{\beta} = \sum_{i=1}^{N}\alpha_i \mathbf{h}(\mathbf{x}_i)^T = \mathbf{H}^T\boldsymbol{\alpha} \tag{4}$$

$$\frac{\partial L_{D_{ELM}}}{\partial \xi_i} = 0 \to \alpha_i = C\xi_i, \quad i = 1,\dots,N \tag{5}$$

$$\frac{\partial L_{D_{ELM}}}{\partial \alpha_i} = 0 \to \mathbf{h}(\mathbf{x})^T\boldsymbol{\beta} - t_i + \xi_i = 0, \quad i = 1,\dots N \tag{6}$$

where $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1),\dots,\mathbf{h}(\mathbf{x}_N)]^T$ is the hidden layer output matrix and $\boldsymbol{\alpha} = [\alpha_1,...,\alpha_N]^T$ is the vector of Lagrange variables. By substituting (4) and (5) into (6), the aforementioned equations can be equivalently written as

$$\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)\boldsymbol{\alpha} = \mathbf{T} \tag{7}$$

here I is the identity matrix and $\mathbf{T} = [t_1,\dots,t_N]^T$. From (4) and (7), we have

$$\boldsymbol{\beta} = \mathbf{H}^T \cdot \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{\mathbf{-1}} \mathbf{T} \tag{8}$$

The output function of ELM classifier (1) can be further derived as

$$f(x) = \mathbf{h}(\mathbf{x})^T\boldsymbol{\beta} = \mathbf{h}(\mathbf{x})^T \cdot \mathbf{H}^T \cdot \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{\mathbf{-1}} \mathbf{T} \tag{9}$$

2.1.3 Application of ELM

Since ELM is able to learn faster and has a better generalization performance, it is very popular and also gets a lot of attention. Many researches have extended the basic ELM to different fields and developed ELM models that focus on different distinguishing features.

Online sequential ELM (OS-ELM) (Liang, Huang, Saratchandran, & Sundararajan, 2006) was a good method to handle time series problems. It could learn sequential data by update the output weights. Research from Zhu et al. (2017) proposed a generalization of ELM which was named Monotonic Classification Extreme Learning Machine (MCELM) in which the monotonicity constraints were imposed on the original ELM. The mathematical model of MCELM was a quadratic programming problem in which both the classification error and monotonicity are taken into account. The experiment results showed that MCELM had a smaller Mean Absolute Error (MAE) value than other methods such as Classification and Regression Tree (CART) (Breiman, Friedman, Olshen, & Stone, 1984), Rank Tree (Xia, Zhang, Li, & Yang, 2008), REMT (Hu et al., 2012), Ordinal Learning Method (OLM) (Ben-David, Sterling, & Pao, 1989), Ordinal Stochastic Dominance Learner (OSDL) (Cao-Van, 2003), and ELM.

Leng et al. (2015) proposed a one-class classifier based on ELM which is constructed for situations where only target class was well described. The concept of SVDD (Tax & Duin, 1999) motivated them to use a decision function for one-class classification problem. For more applications of ELM, Huang, Wang, and Lan (2011) have compiled lots of other ELM variants.

In the past literature, except for the research of Zhu et al. (2017), there are few studies on monotonic classification based on ELM, and there is no ELM related research on one-class and monotonic classification.

2.2 One-class Classification

The purpose of traditional multi-class classification is to classify unknown data into one of several pre-defined categories. For example, binary classification is the simplest case with two categories. An issue occurs when the unknown data is not classified into any categories (Khan & Madden, 2009, 2014). This problem does not occur under the one-class classification. The goal of one-class classification is to distinguish between objects from one class (known as the positive class or target class) and all other possible objects (Tax & Duin, 2001). According to the literature review, we can find that similar conceptual studies of one-class classification have been used in different applications such as novelty detection (Bishop, 1994), and positive-only learning (Parra, Deco, & Miesbach, 1996).

The concept of one-class classification is the phenomenon that patterns from the same class usually cluster regularly together, while patterns from other classes scatter in feature space (Wang, Lopes, & Tax, 2004). In one-class classification, only samples of the target class are required. Wang et al. (2004) proposed a general concept as below

$$Class(x) = \begin{cases} target, & Measurement(x) \le threshold; \\ non\text{-}target, & otherwise \end{cases} \tag{10}$$

We can understand (10) as a distance-based decision function, which can classify a data into target class if the distance is smaller than the threshold, and classify the data as an outlier if the distance is bigger than the threshold.

This project refers to the method proposed by Leng et al. (2015). We build a distance-related decision function outside the ELM model and use the decision function to judge whether the result produced by the monotonic ELM belongs to the target class or not. The establishment and the process of the decision function will be introduced in detail in Section 3.2.2.

2.3 Monotonic Classification

Ordinal classification is a generalization of the traditional classification. In traditional classification problems, the class labels are nominal while in the ordinal classification the class labels are ordered. For example, the purchase desire level after seeing a car can be classified into four levels: slight, medium, strong and very strong. We can find that there is an order relationship between these classes. In other words, ordinal classification employs the information of order among the classes.

A particular case of ordinal classification is monotonic classification. Monotonicity is a common form of domain knowledge imposed by economic theory or human decision-makers (Daniels & Velikova, 2006), such as credit loan approval or investment decision making and evaluation issues. Monotonicity constraints require that the class label assigned to a pattern should be greater or equal than the class labels assigned to the patterns it domains. For example, if the car has a higher level of safety, the customer will have a higher probability of buying the car. Or like the price of a house increases with an increase in the number of rooms or the availability of air conditioning.

Monotonic function means function value will continuously increase (decrease) according to increasing (decreasing) of independent variable value. The interest in the application of monotonic classification has increased over the past few years (Gutiérrez & García, 2016; Zhu et al., 2017). Ben-David et al. (1989) introduced the first algorithm OLM for ordinal classification with monotonic constraints in the machine learning community.

Add monotonic constraints to optimize models can reduce search space (Bonchi, Giannotti, Mazzanti, & Pedreschi, 2003), Sill (1998) proposed monotonic networks in which the monotonicity constraints are enforced by constraining the sign of the hyperplane weight. Lang (2005) proposed a monotonic classifier, monotonic multi-layer perceptron (MonMLP). MonMLP satisfies the monotonicity constraints by constraining the sign of the weights of the multi-layer perceptron network. Monotonicity constraints support vector machine (Chen & Li, 2014) is a rating model based on a support vector machine including monotonicity constraints in the optimization problem, the selected method of monotonicity constraints is randomly generated.

Although monotonic classification has gradually attracted attention in the field of data mining, the past monotonic classification related studies mostly belong to two classes or multi-classes classification (Zhu et al., 2017) (H. Daniels & Velikova, 2010). Namely, there are few studies in one-class classification related to monotonic classification.

2.4 Definition of Monotonic Classification

We use the definition of monotonicity proposed by Cano, Gutiérrez, Krawczyk, Woźniak, and García (2019) here. We formally define a classification dataset with ordinal labels and monotonicity constraints. Assume that patterns are described using a total of n input variables with order domains, $\mathbf{x}_i \subseteq \mathbb{R}^n$, and a class label, y_i, from a finite set of C ordered labels , $y_i \in \mathcal{Y} = \{1,\ldots,C\}$. In this way, the data set D consists of $\mathbf{N}$ samples or instances $D = \{(\mathbf{x}_1, y_1),\ldots,(\mathbf{x}_\mathbf{N}, y_\mathbf{N})\}$. A dominance relation, $\succeq$, is defined as follow:

$$\mathbf{x} \succeq \mathbf{x}' \Leftrightarrow x^{s} \geq x^{s'} \quad \forall s \quad \text{with a monotonicity constraint,} \tag{11}$$

where x^{s} and $x^{s'}$ are the sth coordinates of patterns $\mathbf{x}$ and $\mathbf{x}'$, respectively. By this definition, we can easily tell that $\mathbf{x}$ dominates $\mathbf{x}'$ if each coordinate of $\mathbf{x}$ is not smaller than the respective coordinate of $\mathbf{x}'$.

A pair of comparable examples $(\mathbf{x}, y)$ and $(\mathbf{x}', y')$ is said to be monotone if :

$$\mathbf{x} \succeq \mathbf{x}' \wedge \mathbf{x} \neq \mathbf{x}' \wedge y \geq y' \tag{12}$$

or

$$\mathbf{x} = \mathbf{x}' \wedge y = y' \tag{13}$$

A data set D with n examples is monotone if all possible pairs of examples are either monotone or incomparable.

3. Research Methodology

In this project, we propose a revised ELM model for monotonic one-class classification problem. Section 3.1 defines the problem that we are working on. Section 3.2 contains the methods for satisfying the monotonicity constraints on the network and the decision function for one-class classification. Finally, the revised model that we proposed is described in Section 3.3 and expected experiment and analysis is described in Section 3.4.

3.1 Problem Definition

One-class classification and monotone classification have been widely used in classification applications in the past years. Hence, we propose a classification model that satisfies a one-class and monotonic classification problem. Furthermore, for most of the monotonic classifier, the monotonicity constraints are satisfied by adding constraints in the network structures, and re-construct the model. However, the method we use satisfies the monotonic constraints by revising the weights of the network so that it can be applied in a variety of different network structures.

3.2 Model Construction

3.2.1 Monotonicity Constraints

One way to guarantee an increasing function is to keep all the weights in the neural networks non-negative. Zhang and Zhang (1999) proposed a method to incorporate the known constraints of monotonic relations between inputs and outputs through a new neural network structure. They modified the network weight structure by replacing an ordinary weight in a feedforward network with an exponential form so that all available algorithms can still be applied. In addition, they considered the monotonicity constraints as

$$\frac{\partial y}{\partial x_j} \geq 0 \quad \text{or} \quad \frac{\partial y}{\partial x_j} \leq 0 \tag{14}$$

where $y = f(x_1, x_2, \cdots, x_j, \cdots)$ is an output function of the input variable x_j.

Denote the weighted linear combination of inputs to a neuron j by a_j, the bias b_j, and the output z_j. A single neuron output is defined as

$$z_j = G(a_j) \tag{15}$$

$$a_j = b_j + \sum_i e^{w_{ij}} z_i \tag{16}$$

where G is the activation function and is assumed to be increasing, w_i and e^{w_i} are the ordinary and revising form of weight in neural network. The summation is over all neurons in the previous layer that connect to the current neuron. To show that the output of the network y is always an increasing function of an input x, we consider the partial derivative of y with respect to x, by chain rule

$$\begin{aligned}
\frac{\partial y}{\partial x} &= G'(a) \sum_j e^{w_j} \frac{\partial z_j}{\partial x} \\
\frac{\partial z_j}{\partial x} &= G'_j(a_j) \sum_k e^{w_k} \frac{\partial z_k}{\partial x} \\
&\cdots \\
\frac{\partial z_l}{\partial x} &= G'_l(a_l) \sum_m e^{w_m} \frac{\partial z_m}{\partial x} \\
\frac{\partial z_m}{\partial x} &= G'_m(x)
\end{aligned} \tag{17}$$

We can find that all activation functions are assumed to be incremented, and the exponential functions are positive so that we can get $\partial y / \partial x \geq 0$. We can make a conclusion that y is an increasing function of x.

If the output function is required to be decremented of an input variable, we can apply a preprocessing step

$$p = -x \tag{18}$$

the output function then will be increasing.

3.2.2 Decision Function of One-class ELM

One-class classifier defines a distance function d between the objects and the target class in the training phase. In general, two important parameters are contained in the classification model, threshold θ and model parameter λ. A test sample $\mathbf{z}$ is classified to the target class if $d(\mathbf{z}|X,\lambda)<\theta$, where X is the training set. Leng et al. (2015) proposed a decision function for a one-class ELM classification problem. They used the compactness hypothesis as the basis for object recognition. It states that for similar objects from the target class, the target outputs should be the same, which means:

$$t_i = y, \quad \forall \mathbf{x_i} \in X \tag{19}$$

where y is a real number, and the desired target output vector then is $\mathbf{T}=[t_1,\ldots,t_N]^T=[y,\ldots y]^T$. For a new test sample $\mathbf{z}$, the distance function between the sample objects and the target class is defined as

$$\begin{aligned} d_{ELM}(\mathbf{z}|X,\lambda) &= |\mathbf{h}(\mathbf{z})\boldsymbol{\beta}-y| \\ &= \left|\mathbf{h}(\mathbf{z})\mathbf{H}^T\left(\frac{\mathbf{I}}{C}+\mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T}-y\right| \end{aligned} \tag{20}$$

The distances of the training samples to the target class can be determined using (20) and the constraint of (2)

$$d_{ELM}(\mathbf{x}_i|X,\lambda)=|\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta}-y|=|\xi_i| \tag{21}$$

We can find that the larger $|\xi_i|$ means the more deviant of the training sample $\mathbf{x}_i$ from the target class. Let $d_{(1)}\geq d_{(2)}\geq\ldots\geq d_{(N)}$ stands for the sorted sequence of the distances of training samples, and $\mathbf{d}=[d_{(1)},d_{(2)},\ldots,d_{(N)}]$. The threshold θ can then be written as

$$\theta = d_{floor(\mu\cdot N)} \tag{22}$$

where μ is a user-specified fraction or can be understood as the proportion that the training samples that should be rejected from the target class. $floor(\mu\cdot N)$ is the function that returns the largest integer not greater than $\mu\cdot N$. Afterward, we can get the decision function for one-class ELM classification problem:

$$\begin{aligned} \zeta_{ELM}(\mathrm{z}) &= sign(\theta-d_{ELM}(\mathrm{z}|X,\lambda)) \\ &= \begin{cases} 1 & \text{z is classified as a target} \\ -1 & \text{z is classified as a outlier} \end{cases} \end{aligned} \tag{23}$$

3.3 Model of Monotonic One-class ELM

In this subsection, we present a revised ELM model with monotonic classification and one-class classification materials.

Denote training data and the desired target output by $(\mathrm{x}_i, t_i)_{i=1}^{N}$, N is the number of training data and desired target output, $e^{\boldsymbol{\beta}} = [e^{\beta_1}, e^{\beta_2}, ..., e^{\beta_L}]^T$ is the output weights in exponential form between the hidden layer and the output layer, $e^{\mathbf{w}_j} = [e^{w_{j1}}, e^{w_{j2}}, ..., e^{w_{jn}}]^T$ is the input weights in exponential form that connects input data and j th hidden nodes, b_j is the bias of j th hidden node. One single output node is enough for the one-class classification problem. The monotonic ELM output function can be formulated as

$$f(x) = \mathbf{h}(\mathbf{x})^T e^{\boldsymbol{\beta}} = \sum_{j=1}^{L} e^{\beta_j} G(e^{\mathbf{w}_j}, b_j, \mathbf{x}) \tag{24}$$

where $\mathbf{h(x)}$ is the output vector of the hidden layer corresponding to the input data $\mathbf{x}$. $\mathbf{h(x)}$ maps the data from N dimensional input space to the L dimensional hidden layer space (ELM feature space) (Huang, Zhou, Ding, & Zhang, 2011). $G(e^{\mathbf{w}}, b, \mathbf{x})$ is the activation function which satisfies ELM universal approximation capability theorem (Huang, Chen, & Siew, 2006). As mention in section 2, the objective function of monotonic one-class ELM can be formulated as

$$\begin{aligned} \min\ L_{P_{Mono-ELM}} &= \frac{1}{2}\left\|e^{\beta}\right\|^2 + \frac{C}{2}\sum_{i=1}^{N}\left\|\xi_i\right\|^2 \\ \text{s.t.}\ \ \mathrm{h}(\mathrm{x}_i)^T e^{\beta} &= t_i - \xi_i, \qquad i = 1, ..., N. \end{aligned} \tag{25}$$

Based on the KKT theorem, to train ELM is equivalent to solve the following dual optimization problem:

$$L_{D_{Mono-ELM}} = \frac{1}{2}\left\|e^{\boldsymbol{\beta}}\right\|^2 + \frac{C}{2}\sum_{i=1}^{N}\left\|\xi_i\right\|^2 - \sum_{i=1}^{N}\alpha_i(\boldsymbol{h}(\boldsymbol{x}_i)^T e^{\boldsymbol{\beta}} - t_i + \xi_i) \tag{26}$$

each Lagrange multiplier α_i corresponds to the i th training data. The KKT optimality conditions are shown below:

$$\frac{\partial L_{D_{Mono-ELM}}}{\partial \boldsymbol{\beta}} = 0 \to \boldsymbol{\beta} = \ln(\sum_{i=1}^{N}\alpha_i \mathbf{h}(\mathbf{x}_i)^T) \to \boldsymbol{\beta} = \ln(\mathbf{H}^T\boldsymbol{\alpha}) \tag{27}$$

$$\frac{\partial L_{D_{Mono-ELM}}}{\partial \xi_i} = 0 \to \alpha_i = C\xi_i, \qquad i = 1, ..., N \tag{28}$$

$$\frac{\partial L_{D_{Mono-ELM}}}{\partial \alpha_i} = 0 \to \mathbf{h}(\mathbf{x}_i)^T e^{\boldsymbol{\beta}} - t_i + \xi_i = 0, \qquad i = 1, ..., N \tag{29}$$

where $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1), \ldots, \mathbf{h}(\mathbf{x}_N)]^T$ is the hidden layer output matrix and $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]^T$ is the vector of Lagrange variables. By substituting (27) and (28) into (29) we can get the equation of (29) equivalently written as

$$\left(\frac{\mathbf{I}}{C}+\mathbf{H}\mathbf{H}^T\right)\boldsymbol{\alpha}=\mathbf{T} \tag{30}$$

where I is the identity matrix and $\mathbf{T}=[t_1,\ldots,t_N]^T$. From (27) and (30), we have

$$\boldsymbol{\beta}=\ln\left(\mathbf{H}^T\cdot\left(\frac{\mathbf{I}}{C}+\mathbf{H}\mathbf{H}^T\right)^{\mathbf{-1}}\mathbf{T}\right) \tag{31}$$

the output function of monotonic ELM can be written as

$$f(x)=h(x)^T e^{\boldsymbol{\beta}}=h(x)^T\cdot\mathbf{H}^T\cdot\left(\frac{\mathbf{I}}{C}+\mathbf{H}\mathbf{H}^T\right)^{\mathbf{-1}}\mathbf{T} \tag{32}$$

3.4 Experiment and Analysis

We plan to conduct empirical experiments to validate the effectiveness of the research methods to be developed in this project. The research methods will be tested on the data sets that are the most commonly used in the monotonic classification literatures (Cano, Gutiérrez, Krawczyk, Woźniak, and García, 2019). The data sets include Auto MPG, Boston Housing, Car, ERA, ESL, LEV, Machine CPU, Pima, and SWD (Cano, Gutiérrez, Krawczyk, Woźniak, and García, 2019).

4. Project Anticipations

4.1 Anticipated Accomplishments

The objectives of this project will be accomplished in one year and we will complete the following tasks:

1. Comprehensively analyzing the researches that related to this project, such as Extreme Learning Machine (ELM), Monotonic Classification Extreme Learning Machine (MCELM), one-class classification and monotonic classification.
2. Theoretically formulating the revised ELM model for monotonic classification.
3. Design a Monotonic One-class Extreme Learning Machine framework.
4. Developing algorithm and programs for the Monotonic One-class Extreme Learning Machine framework.
5. Conducting the experiments to investigate the framework with real datasets, and comparing the performance and results.
6. Documenting the project results and submitting to international conferences or journals.

4.2 Anticipated Contributions

In prior literatures, there have been few studies that work on monotonic classification based on ELM network structure. The proposed monotonic classification ELM in the literature satisfied the monotonicity constraints by imposing the constraints to the original ELM, however, the structure of the original ELM model then had been changed. The objective of this project is to maintain the original ELM structure, aim to design a revised ELM framework for the monotonicity constraints. We satisfy the monotonicity by adjusting the weights into exponential forms, so we can maintain the original network structure of ELM and achieve the goal of monotonic classification. Furthermore, we establish a distance-related decision function to classify one-class data. Then, a monotonic one-class classifier model based on ELM will be constructed.

The novelty of this project is the construction of the monotonicity constraints for the original ELM. Different from the proposed model in prior literature, our model is much more robust since it adjusts the weights into exponential forms to achieve the goal of monotonic classification, and not to change the original network structure of ELM. Furthermore, the model will be evaluated with a distance-related decision function to classify one-class data. In addition, research assistants have the opportunities to learn how to search and trace related publications from research publication database. Developing new algorithms for the proposed monotonic one-class classifier model based on ELM, which is different from previous studies, brings the participants independent thinking skills that are beneficial for their future academic researches. The conduction process for the overall project involves activities such as problem formulation, algorithm development, programming, and project management provide participants empirical experiences for their career development. Additionally, the experimental analysis enables their ability to conduct experiments for assessing the proposed methodology.

4.3 Anticipated Difficulties and Solutions

The tasks to be conducted are illustrated in Section 4.1 involve theoretical modeling, designing of algorithms, experimentation and analysis, model development and implementations of real-world datasets. The underlying theories that will be used in this project include Extreme Learning Machine (ELM), monotonic classification, one-class classification, algorithm design, which will be difficult for fully understanding and applications. In order to accomplish the objectives of this project, the anticipated difficulties and solutions are described as follows.

1. A senior research assistant of PhD student will join this project for reviewing related literatures, surveying relevant theories, developing theoretical framework, designing the experiments, analyzing the experimental results, organizing the reports, and so on. Moreover, by helping master students understand the underlying theories and discussion, the PhD student have to consolidate the research processed based on the theory and help to evaluate the feasibility of the framework.
2. Three research assistants of master student to aid developing algorithm, conducting experimentation and analysis, designing the solution systems.
3. The computational complexity is expected to be high during the training phase of the monotonic one-class classifier model based on ELM. To accelerate the training process, a GPU-based server will be used for model construction. Benefit from the high-performance computer facilities, more experiments will be conducted to find more stable and robust prediction model for this project.

References

Bartlett, P. L. (1998). The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE transactions on Information Theory, 44*(2), 525-536.

Ben-David, A., Sterling, L., & Pao, Y. H. (1989). Learning and classification of monotonic ordinal concepts. *Computational Intelligence, 5*(1), 45-49.

Bishop, C. M. (1994). Novelty detection and neural network validation. *IEE Proceedings-Vision, Image and Signal processing, 141*(4), 217-222.

Bonchi, F., Giannotti, F., Mazzanti, A., & Pedreschi, D. (2003). *ExAMiner: Optimized level-wise frequent pattern mining with monotone constraints.* Paper presented at the Third IEEE International Conference on Data Mining.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees.

Cano, J.-R., Gutiérrez, P. A., Krawczyk, B., Woźniak, M., & García, S. (2019). Monotonic classification: An overview on algorithms, performance measures and data sets. *Neurocomputing, 341*, 168-182.

Cao-Van, K. (2003). *Supervised ranking: from semantics to algorithms.* Ghent University,

Chen, C.-C., & Li, S.-T. (2014). Credit rating with a monotonicity-constrained support vector machine model. *Expert Systems with Applications, 41*(16), 7235-7247.

Daniels, H., & Velikova, M. (2010). Monotone and partially monotone neural networks. *IEEE Transactions on neural networks, 21*(6), 906-917.

Daniels, H. A., & Velikova, M. V. (2006). Derivation of monotone decision models from noisy data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 36*(5), 705-710.

Gutiérrez, P. A., & García, S. (2016). Current prospects on ordinal and monotonic classification. *Progress in Artificial Intelligence, 5*(3), 171-179.

Hu, Q., Che, X., Zhang, L., Zhang, D., Guo, M., & Yu, D. (2012). Rank Entropy-Based Decision Trees for Monotonic Classification. *IEEE Transactions on Knowledge and Data Engineering, 11*(24), 2052-2064.

Huang, G.-B., & Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing, 70*(16-18), 3056-3062.

Huang, G.-B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks, 17*(4), 879-892.

Huang, G.-B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. *International journal of machine learning and cybernetics, 2*(2), 107-122.

Huang, G.-B., Zhou, H., Ding, X., & Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42*(2), 513-529.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. *Neural networks, 2*, 985-990.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing, 70*(1-3), 489-501.

Khan, S. S., & Madden, M. G. (2009). *A survey of recent trends in one class classification.* Paper presented at the Irish conference on artificial intelligence and cognitive science.

Khan, S. S., & Madden, M. G. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review, 29*(3), 345-374.

Lang, B. (2005). *Monotonic multi-layer perceptron networks as universal approximators.* Paper presented at the International conference on artificial neural networks.

Leng, Q., Qi, H., Miao, J., Zhu, W., & Su, G. (2015). One-class classification with extreme learning machine. *Mathematical problems in engineering, 2015*.

Li, K.-L., Huang, H.-K., Tian, S.-F., & Xu, W. (2003). *Improving one-class SVM for anomaly detection.* Paper presented at the Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693).

Liang, N.-Y., Huang, G.-B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks, 17*(6), 1411-1423.

Parra, L., Deco, G., & Miesbach, S. (1996). Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation, 8*(2), 260-269.

Perdisci, R., Gu, G., & Lee, W. (2006). *Using an Ensemble of One-Class SVM Classifiers to Harden Payload-based Anomaly Detection Systems.* Paper presented at the ICDM.

Shin, H. J., Eom, D.-H., & Kim, S.-S. (2005). One-class support vector machines—an application in machine fault detection and classification. *Computers & Industrial Engineering, 48*(2), 395-408.

Sill, J. (1998). *Monotonic networks.* Paper presented at the Advances in neural information processing systems.

Tax, D. M., & Duin, R. P. (1999). Support vector domain description. *Pattern Recognition Letters, 20*(11-13), 1191-1199.

Tax, D. M., & Duin, R. P. (2001). Uniform object generation for optimizing one-class classifiers. *Journal of machine learning research, 2*(Dec), 155-173.

Tax, D. M. J. (2002). One-class classification: Concept learning in the absence of counter-examples.

Tissera, M. D., & McDonnell, M. D. (2016). Deep extreme learning machines: supervised autoencoding architecture for classification. *Neurocomputing, 174*, 42-49.

Wang, Q., Lopes, L. S., & Tax, D. M. (2004). *Visual object recognition through one-class learning.* Paper presented at the International Conference Image Analysis and Recognition.

Xia, F., Zhang, W., Li, F., & Yang, Y. (2008). Ranking with decision tree. *Knowledge and information systems, 17*(3), 381-395.

Zhang, H., & Zhang, Z. (1999). *Feedforward networks with monotone constraints.* Paper presented at the IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339).

Zhu, H., Tsang, E. C., Wang, X.-Z., & Ashfaq, R. A. R. (2017). Monotonic classification extreme learning machine. *Neurocomputing, 225*, 205-213.

五、申請補助經費：

（一）請將本計畫申請書之第七項(表CM07)、第八項(表CM08)、第九項(表CM09)、第十項(表CM10)、第十一項(表CM11)、第十二項(表CM12)所列費用個別加總後，分別填入「研究人力費」、「耗材、物品、圖書及雜項費用」、「國外學者來臺費用」、「研究設備費」、「國外差旅費-執行國際合作與移地研究」及「國外差旅費-出席國際學術會議」等欄內。
（二）管理費為申請機構配合執行本計畫所需之費用，其計算方式係依本部規定核給補助管理費之項目費用總和及各申請機構管理費補助比例計算後直接產生，計畫主持人不須填寫「管理費」欄。
（三）「貴重儀器中心使用額度」係將第十三項(表CM13)所列使用費用合計數填入。
（四）請依各年度申請博士級研究人員之名額填入下表，如於申請時一併提出「補助延攬博士級研究人員(含大陸)員額/人才進用申請書」（表CIF2101、CIF2102），若計畫核定僅核定名額者應於提出合適人選後，另依據本部「補助延攬客座科技人才作業要點」規定向本部提出進用申請，經審查通過後，始得進用該名博士級研究人員。
（五）申請機構或其他單位（含產業界）提供之配合項目，請檢附相關證明文件。

金額單位：新臺幣元

補助項目 ＼ 執行年次		第一年（109年8月～110年7月）	第二年	第三年	第四年	第五年
業務費		**719,000**				
研究人力費		624,000				
耗材、物品、圖書及雜項費用		95,000				
國外學者來臺費用		0				
研究設備費		**35,000**				
國外差旅費		**0**				
執行國際合作與移地研究		0				
出席國際學術會議		0				
管理費		**113,100**				
合計		**867,100**				
貴重儀器中心使用額度		0				
博士級研究人員	國內、外地區	共 0 名	共 名	共 名	共 名	共 名
	大陸地區	共 0 名	共 名	共 名	共 名	共 名

申請機構或其他單位（含產業界）提供之配合項目（無配合補助項目者免填）

配合單位名稱	配合補助項目	配合補助金額	配合年次	證明文件

六、主要研究人力：

（一）請依照「主持人」、「共同主持人」、「協同研究人員」及「博士級研究人員」等類別之順序分別填寫。

類 別	姓名	服務機構/系所	職稱	在本研究計畫內擔任之具體工作性質、項目及範圍	＊每週平均投入工作時數比率(%)
主持人	李昇暾	國立成功大學工業與資訊管理學系（所）	教授	主導研究方向、方法的實施、模型建置與實驗測試分析等工作。	30%

※ 註：每週平均投入工作時數比率係填寫每人每週平均投入本計畫工作時數佔其每週全部工作時間之比率，以百分比表示（例如：50%即表示該研究人員每週投入本計畫研究工作之時數佔其每週全部工時之百分五十）。

（二）如申請博士級研究人員，請另填表CIF2101及CIF2102(若已有人選者，請務必填註人選姓名，並將其個人資料表(表C301～表C303)併同本計畫書送本部)。

七、研究人力費：

（一）凡執行計畫所需研究人力費用，均得依本部「補助專題研究計畫研究人力約用注意事項」規定，按所屬機構自訂敘薪標準及職銜，就預估專任、兼任人員或臨時工需求填寫，並請述明該研究人力在本計畫內擔任之具體內容、性質、項目及範圍，以利審查。專任人員不限學歷，包含博士級人員。

（二）約用專任人員，請依其於專題研究計畫負責之工作內容，所應具備之專業技能、獨立作業能力、預期績效表現及相關學經歷年資等條件，綜合考量敘薪，並檢附各機構自訂之薪資支給依據，以為本部核定聘用助理經費之參考。

（三）請分年列述。

第 1 年　　　　金額單位：新臺幣元

類別	金額	請敘明在本計畫內擔任之具體內容、性質、項目及範圍 (如約用專任人員，請簡述其於計畫內所應具備之專業技能、獨立作業能力、預期績效表現及相關學經歷年資等條件)
兼任人員(碩士生-學習範疇)	216,000	文獻蒐集與研讀、蒐集實驗資料、資料前處理、程式撰寫、系統整合與測試。 (月支費用 9000.00元 x 12.00月) x 2名
兼任人員(博士候選人-學習範疇)	408,000	文獻蒐集與研讀、模式理論推導與建立、蒐集實驗資料、程式開發撰寫、系統整合與測試、結果分析。 (月支費用 34000.00元 x 12.00月) x 1名
合計	624,000	

八、耗材、物品、圖書及雜項費用：

(一) 凡執行研究計畫所需之耗材、物品(非屬研究設備者)、圖書及雜項費用，均可填入本表內。
(二) 說明欄請就該項目之規格、用途等相關資料詳細填寫，以利審查。
(三) 若申請單位有配合款，請於備註欄註明。
(四) 請分年列述。

第 1 年　　　　金額單位：新臺幣元

項 目 名 稱	說明	單位	數量	單價	金額	備註
消耗性器材	雷射印表機碳粉匣	支	4	4,000	16,000	
論文發表費	投稿研究成果與發表論文費用	篇	3	6,000	18,000	
雜支	隨身碟、空白光碟片與文具用品等雜項支出、研究資料影印費、影印紙	式	1	17,000	17,000	
電腦使用費	維修科技部資助購買之電腦與印表機等週邊設備及其所須零件	次	2	5,000	10,000	
郵電費	聯絡計劃相關事宜費用	月	12	500	6,000	
差旅費	資料蒐集與專家學者訪談費用	次	4	4,000	16,000	
英文論文修改費	投稿國際英文期刊論文修改費用	篇	2	6,000	12,000	
合　　計					95,000	

十、研究設備費：

（一）凡執行研究計畫所需單價在新台幣一萬元以上且使用年限在二年以上與研究計畫直接有關之各項設備屬之。各類研究設備金額請於金額欄內分別列出小計金額。
（二）購置設備單價在新臺幣二十萬元以上者，須檢附估價單。
（三）若申請機構及其他機構有提供配合款，請務必註明提供配合款之機構及金額。
（四）儀器設備單價超過六十萬元(含)以上者，請詳述本項設備之規格與功能(諸如靈敏度、精確度…等)，其他重要特性與重要附件，以及申購本設備對計畫執行之必要性。本項設備若獲補助，主持人應負維護保養之責，並且在不妨礙個人研究計畫或研究群計畫之工作下，同意提供他人共同使用，以避免設備閒置。
（五）計畫主持人執行本項研究計畫，如欲申請購置單價新臺幣一千萬元(含)以上之大型儀器，請填表CM10-1。該項設備若獲本部核定補助新臺幣一千萬元(含)以上，則單獨核給一個規劃計畫，主持人須遵守本部大型儀器之管考規定。
（六）經本部補助之大型儀器，儀器資訊須公開於本部全球資訊網之「貴重儀器開放共同管理平台」。
（七）請分年列述。

第 1 年　　　　金額單位：新臺幣元

類別	設備名稱（中文/英文）	說明	數量	單價	金額	經費來源	
						本部補助經費需求	提供配合款之機構名稱及金額
儀器及資訊設備	高效能個人電腦主機/High-performance Personal computer	開發建置系統之用，因需針對複雜之二次數學規畫法求解，需高效能運算之主機	1	35,000	35,000	35,000	
合計					35,000	35,000	

十四、近三年內執行之研究計畫

（請務必填寫近三年所有研究計畫，不限執行本部計畫）

計畫名稱 （本部補助者請註明編號）	計畫內擔任之工作	起迄年月	補助或委託機構	執行情形	經費總額
不同能耐生命週期下策略決策支援系統之建立(108-2410-H-006-108-MY2)	共同主持人	2019/08/01～2021/07/31	科技部	執行中	2,726,000
深度強化學習法於金融科技時間序列預測之研究(108-2410-H-006-106-MY3)	主持人	2019/08/01～2022/07/31	科技部	執行中	2,548,000
深度單類別文本分類法於偵測藥物不良反應之研究(108-2410-H-006-107-)	主持人	2019/08/01～2020/07/31	科技部	執行中	808,000
智能鞋墊虛實整合之創新服務營運模式(108-2745-8-006-004-)	主持人	2019/06/01～2020/05/31	科技部	執行中	917,000
不同能耐生命週期下策略決策支援系統之建立(107-2410-H-006-043-)	共同主持人	2018/08/01～2019/07/31	科技部	已結案	1,088,000
隱性主題模型於群眾募資預測之研究(107-2410-H-006-046-)	主持人	2018/08/01～2019/10/31	科技部	已結案	816,000
不同能耐生命週期下策略決策支援系統之建立(106-2410-H-006-008-)	共同主持人	2017/08/01～2018/07/31	科技部	已結案	1,235,000
TOPSIS群體決策環境下整合策略之研究(105-2410-H-006-037-)	主持人	2016/08/01～2017/10/31	科技部	已結案	900,000
具先前領域知識大型支援向量機之研究(105-2410-H-006-038-MY3)	主持人	2016/08/01～2019/10/31	科技部	已結案	2,821,000
建構以文字探勘為基礎之感性工學系統於顧客需求分析之研究(105-2410-H-143-020-)	共同主持人	2016/08/01～2017/10/31	科技部	已結案	604,000
競爭策略決策支援系統(103-2410-H-006-054-MY3)	共同主持人	2014/08/01～2017/07/31	科技部	已結案	3,278,000
健保門診申報資料診斷編碼品質評估計畫	協同主持人	2018/12/29～2019/06/10	衛福部健保署	已結案	0
合　　計					17,741,000

科技部人文司專題研究計畫主持人研究績效表(2017.09.06 製表)

計畫主持人：李昇暾　　　　職稱：特聘教授
服務機關：國立成功大學工資管系

1. Monotonically data mining

Traditional data mining research mainly focuses on developing, demonstrating, and pushing the use of specific algorithms and models based on the data collected from observations. However, when solving real-world problems using data mining, there is usually prior knowledge that exists in addition to the collected data. For example, in loan-approval applications, if loan applicants A and B have the same attribute values, except that applicant A has a higher income than applicant B, then applicant A has a greater chance of getting the loan than applicant B. In other words, "a higher income increases the chance of loan approval" is an example of prior domain knowledge. Incorporating this kind of background knowledge into data mining is on the list of "challenging problems in data mining research" and is acknowledged as a difficult problem.

Since 2010, we have devoted to the study of monotonic data mining in the framework of well-recognized Support Vector Machines and proposed two good-reputation models as monotonicity-constrained SVM (MC-SVM) and MC-FSVM (fuzzy SVM), which is published in IEEE Transactions on Fuzzy Systems, the top one journal of fuzzy logic. We continue to extend and enhance our work to tackle various important issues in the area. For example, we adapted MC-SVM to the regression problem (MC-SVR). The resulting paper was just accepted by Neural Computing and Applications, Dec. 2019. In addition, we move forward to deal with how to select effect and efficient monotonic constraints, and how to deal with the scalability of big data using Hadoop cloud computing and the streaming data using incremental learning. These outcomes are to be submitted to *Knowledge-Based Systems* and/or *Neurocomputing* journals. Currently, we are investigating how to incorporating monotonicity prior knowledge into the context of one-classification classification, which is, to our best knowledge, the first work in the literature. The experimental results are quite encouraging and a draft has been completed and is under revising. We will submit it shortly to top ranked journals.

More importantly, due to our contribution recognized by the area, a Ph.D. student from University of Granada, Spain, recommended by his advisor, came to join our team two years ago under supports by European office of MOST. This international cooperation leads to two research papers on two excellent journals in the area; one was published in Information Sciences (SCI) in 2019 and the other was just accepted by Neurocomputing (SCI), Dec. 2019.

2. Time-series forecasting

Traditional time series forecasting such as statistics and neural networks are usually extensively dependent on historical data, which can be incomplete, imprecise and ambiguous. These uncertainties could be widespread in real-world data and hinder forecasting accuracy, thus limiting the applicability of forecasting models. Unlike traditional time series forecasting approaches, we turned to the fuzzy approach, capable of dealing with vague and incomplete time series data under uncertain circumstances. Since 2007, we have proposed several models to improve the forecasting performance or extend the functionality of dealing problems. One representative work is that we theoretically and experimentally solved the major hurdle of choosing an appropriate *k*-order in high-order forecasting models for fuzzy time series. Other works includes the pioneering work of applying probabilistic approach, hidden Markov Model, in contrast to the traditional rule-based approach used in most of literature to establish the fuzzy logic relationships existing in a fuzzy time series. In addition, we also extended the current fuzzy time series to solve the long-term multiple-step ahead forecasting problem, which has been ignored in the area. These outcomes appeared in the top journals in the area including IEEE Transactions on Fuzzy Systems, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, and Fuzzy Sets and Systems.

3. Text-mining and sentiment analysis

In the past years, we mainly applied text mining to solve some real-world managerial problems. For example, in the work of “*Does reliable information matter? Towards a trustworthy co-created recommendation model by mining unboxing reviews*,” which targeted the problem of information overload due to the huge amount of reviews posted daily which complicates the efforts of consumers to locate reliable information when making a purchase decision. We mined unboxing reviews from Mobile01, calculated the trust scores of the reviewers, and then generated the recommended products by combing this information with customer preferences using a multi-criteria decision-making method. This work was the first study in the literature to explore the popular phenomenon of unboxing forum in Taiwan. The outcome received highly positive recognition and published in *Information Systems and e-Business Management.*

Another related and recent work is “*Do our words or social relations matter to our reputation? Locating reputable reviewers via linguistic styles*.” In this study, we investigate the important issue of locating reputable reviewers in opinion-sharing communities. Numerous studies examining this using social relations have noted the possibility of manipulation. In contrast, we investigate language use, which differs from person to person, and develop a novel prediction model to alleviate such problems. We identify four stylistic aspects and explore their impacts on a reviewer’s reputation. The analyses show that the proposed model can more accurately locate reputable reviewers than the baseline model. In addition, reviewers’ words matter more than social relations do, although a combination of these will boost reviewers’ reputations to a greater extent than one alone. This outcome was published recently by *Information & Management* (SSCI).

4. Group decision-making

In this research area, we focused on developing an effective GMCDM (group multi-criteria decision making) mechanism to allow aggregation and consideration of numerous (often conflicting) criteria in order to choose, rank, sort or describe a set of alternatives to aid a decision process. In particular, we proposed a novel TOPSIS-based GMCDM method to derive a power planning model for Chunghwa Telecom telecom rooms utilizing the senior staff's tacit knowledge. The novelty of proposed method includes an attitude factor, which indicates the degree of strictness of decision makers' attitudes toward a decision problem, in order to identify the superiority and inferiority of each alternative with regard to the others. More importantly, a new aggregation strategy is proposed on the basis of the separation ratio that is used to measure the differences among decision makers. The resulting model provides telecom room power resources planners and managers with technical documents that can help in their decision-making. This work was published in *Omega,* 2014. A real-world case application of the above-mentioned model was published in Sun Yat-Sen Management Review (TSSCI), 2018.

Another GDM work is about "*Using mutually validated memories of experts for case-based knowledge systems,*" published in *Knowledge-Based Systems*, 2015. This work proposed an important index, knowledge consistency, characterized by measures of both individual and group consistencies, can provide a more effective assessment to assign suitable experts' weights than most existing GDM models in the aggregation process, and thus help to reach a more appropriate and realistic group consensus.

5. Medical informatics

Upon the coming of big data era, the corporation of interdisciplinary becomes more important and popular. Based on my experience of data mining, text mining, and big data analytics, I was invited to participate some interesting medical projects in public health to help the research team more effectively and efficiently utilizing the medical big data. The issues under investigation are more related to decision- and strategy-making in government and management. For example, in the work of *"An analysis of the current status and utilization of government health and welfare open data in Taiwan,"* we examined the current status and utilization of government health and welfare open data in Taiwan using characteristic information from datasets released by the Ministry of Health and Welfare and related Institutions available at data.gov.tw. The important finding is that although the number of datasets released by health institutions was lower than that of other sectors, the number of applications using government health and welfare open data was relatively high. Some of the datasets were divided into too many small datasets by some health institutions however, and this did not follow the primacy principle.

Another more interesting work which successfully integrates managerial (particularly MIS) and medical expertise is "*When high-volume PCI operators in high-volume hospitals move to lower volume hospitals—Do they still maintain high volume and quality of* *outcomes?*" published in *Catheterization and Cardiovascular Interventions*, 2018 (SCI, IF: 2.602). Based on the theory established by management science researchers: if star knowledge workers' performance can migrate across organizations, we examined whether high-volume percutaneous coronary intervention (PCI) operators still maintain high volume and quality of outcomes when they moved to lower volume hospitals. The dataset used is from Taiwan National Health Insurance claims data 2000–2012 to identify 98 high-volume PCI operators, 10 of whom moved from one hospital to another during the study period. The analytic result indicates that high-volume operators cannot maintain high volume when they moved from high to moderate or low-volume hospitals; however, the quality of care is maintained. High PCI volume and high-quality outcomes are less portable and more hospital bound. This finding is greatly helpful for hospital managers in strategy making of human resource management.

Another related work of incorporating my expertise in BI and visualization was published in Circulation: Cardiovascular Quality and Outcomes (SCI), 2017 with the title: Graphical Representations of Mortality Data with Confidence Intervals. We proposed a novel visualization approach with statistical analysis contributed to a new way to investigate the morality data of USA.

二、請簡述上述代表性研究成果之個人重要貢獻。

The aforementioned research outcomes accomplished in the past five-ten years are mainly performed by the master and Ph.D. graduate students under my supervision. I have been enjoyed to cook interesting research ideas via reading large amount of literatures, weekly group discussion, participating international conferences and workshops, serving for paper reviewers, etc. I also enjoy in various activities during academic papers submission including writing, revising, polishing, defending, and even proofreading. Meanwhile, I like to contribute my expertise for inter-disciplinary study like medical informatics to make the cooperation work successful.

三、請簡述個人近 5 年對人才培育、研究團隊及學術社群之建立與服務，以及對國家、社會、經濟發展之重要貢獻。

In the past 5 years, there are 20 master and 4 Ph.D. students graduated, respectively. Most master graduates are working for big high-tech companies such as TSMC, TradeVan, Innolux, etc. Two of the four Ph.D. graduates are as faculty members in private and national universities, both are recently promoted to associate professors. Another graduate has been worked for Chunghwa Telecom more than 10 years as a senior engineer. Our team contributing to the academic and industrial societies in cultivating critical skills in big data analytics, data mining, text mining, decision making, medical informatics, etc. In addition, we also contribute the important impacts to the nation, economic development and society summarized as follows.

1. Monotonic data mining. The outcome of monotonic data mining has attracted Innolux Corporation's invitation to have an industrial-academia project under review by MOST, Dec. 2019.

2. Time-series forecasting. The problem of time series forecasting plays an important role in various domains, such as air pollution, population growth, rainfall prediction, and stock forecasting. An industrial-academia project with Bank SinoPac is under review by NCKU, Dec. 2019.

3. Text mining and sentimental analysis. The unboxing work provides academics and practitioners with further insights into the advantages of combining reviewers' trust scores, reviews and consumer preferences when attempting to develop a recommendation model. From a consumer perspective, the proposed model can provide credible information from active reviewers, such as the product score and sentiment orientation of each recommendation, as well as the trust scores of the reviewers, and can thus help when making purchasing decisions. As for manufacturers, organizational managers can co-create value with customers by reading reviews from a large number of users, and this would enable them to gain significant insights into market trends, as well as how to decrease overhead costs.

The work of locating reputable reviewers has significant managerial implications that includes enables potential customers to quickly find reliable information to make purchase decisions, while avoiding reviews written by less reputable reviewers. It also can help manufacturers and retailers find reputable reviewers who can help spread product information rapidly in opinion-sharing communities, thus supporting marketing campaigns. Firms could therefore adopt marketing strategies based on influential reviewers. With regard to reviewers, understanding which stylistic features influence their reputations can help them adjust their writing style.

4. Group decision making. Group decision making is one of the most important activities in business, and high quality decisions rely on good information and experience. Our work definitely fulfills the growing need to help ICT companies like CHT derive a more efficient and systematic allocation strategy to power resource planning. Therefore, it can be served as the strong enabler to facilitate a more effective and efficient resource management in various businesses and industries.

5. Medical informatics. Based on our inter-disciplinary work as afore-mentioned, we suggest the government that the quality and quantity of value-added analyses should be improved and that feedback be provided to the institutions releasing the datasets in order to establish a good value added analyses ecosystem. Regarding to the work of performance migration in medical institutions, it can be used as a guideline to help institutions establish a

more effective strategy in human resource management.