Application of machine learning techniques to predict entrepreneurial firm valuation

Ruling Zhang
Zengrui Tian
Killian J. McCarthy
Xiao Wang
Kun Zhang

Venture capital (VC) is the main contributor to entrepreneurial firms' funding and thus plays a crucial role in their sustainable development and rapid growth. However, early-stage VC investors often face valuation obstacles to predict firm valuation since entrepreneurial firms lack operational performance records and information asymmetry exists between them. In this paper, an integrated differential evolution algorithm and adaptive moment estimation method scheme (Adam-ENN) is proposed for early-stage VC investors to predict entrepreneurial firm valuation. Experimental results show that the proposed machine learning method outperforms the baseline methods. The feature contribution analysis and partial dependence plots were performed to open up the black box of the relationships between entrepreneurial firm valuation and its features. Results indicate that the number of VC investors in the funding syndicate is the most important feature, and VC investors' social capital also plays a significant role in the prediction model. Interestingly, the number of patents cannot convey an effective signal in entrepreneurial firm quality especially in the early-stage entrepreneurial firm valuation. Finally, this paper helps to guide entrepreneurial firms' valuation using machine learning techniques and offers deep insight into the entrepreneurship financing mechanism from the perspective of VC.

KEYWORDS
entrepreneurship financing, investment decision-making, machine learning, valuation prediction, venture capital

1 INTRODUCTION

Entrepreneurial firms are increasingly playing a crucial role in the technological innovation and economic development of the country in recent decades (Sharatchilev et al., 2018). However, only approximately 60% of them can survive for more than 3 years according to a recent research investigation (Hyytinen et al., 2015). The absence of funding has become one of the key barriers for entrepreneurial firms to sustain their development. In this case, venture capital (VC) investors are devoted to addressing entrepreneurial firms' resource and financial constraints, which hence promotes the long-term growth and prosperity of the regional and national economics and protects the entrepreneurship and environment (Bocken, 2015; Ren et al., 2020; Wu et al., 2020). Put another way, VC is the key source of financing and the main contributor to entrepreneurial firms' sustainable development (Gompers & Lerner, 2001; Wang et al., 2015).
Valuation occurs naturally when VC investors decide to make investments (Gompers et al., 2020). Reasonable valuation of an entrepreneurial firm is necessary since overvalued entrepreneurial firms will increase the investment risks of VC investors, whereas under-valued one is likely to conclude negotiations with VC investors, which may lead to the loss of good investment opportunity. However, it is rather difficult for VC investors to evaluate the value of entrepreneurial firms in the early-stage financing rounds. Given that early-stage entrepreneurial firms generally have not yielded profits and lack historical records, their value cannot be estimated by typical and common-used valuation measures such as internal rate of return (IRR).1 Indeed, almost 50% of early-stage VC investors make gut investment decisions by adopting qualitative methods such as intuition (Gompers et al., 2020). Under this circumstance, information asymmetries and market uncertainty drive VC investors to rely on signals outside to judge the quality and potential of ventures. Therefore, it is vital for us to figure out the signals (features) that impact the entrepreneurial firm’s valuation and construct the valuation model for the early-stage VC investors to assist them to make investment decisions.

Entrepreneurial firm valuation is receiving ever more attention from scholars and practitioners (Tumasjan et al., 2021). Yet further development is largely hindered by methodological challenges. Specifically, previous research mainly focuses on the impact of one aspect (only a few features) on firm valuation via traditional research methods (i.e., linear regression or qualitative method). For example, Alexy et al. (2012) mainly investigate how the social capital of VC investors affects entrepreneurship financing using regression techniques, while Gompers et al. (2020) explore the role of entrepreneurial firm valuation playing in entrepreneurship financing through a qualitative framework. Consequently, the unified valuation framework regarding entrepreneurial firm valuation in the early-stage financing rounds has not been constructed, and the existing classical research method still lacks good generality owing to the weak ability to deal with complex, wide, and noisy data (Ponta et al., 2022). In recent years, machine learning methods especially neural network have shown tremendous promise in solving complicated nonlinear problems (Rebala et al., 2019). In fact, there have been few rigorous studies to investigate how to get a comprehensive entrepreneurial firm valuation model for early-stage VC investors taking advantage of machine learning algorithms.

Moreover, the interpretability problem of results is a well-acknowledged problem associated with the application of machine learning models (Rebala et al., 2019). The difficulty of interpreting the results hampers unraveling the impact mechanism between input features and the output variable. Recently, feature contribution analysis and partial dependence plot are proposed to address the aforementioned issues and have been proven to be effective in solving model interpretability problems (Molnar et al., 2021; Ponta et al., 2022). Yet the applications of partial dependence techniques to entrepreneurship financing problems such as entrepreneurial firm valuation adopted by VC investors are still very limited.

In order to fulfill the aforementioned research gaps, the research tasks of this article are to (1) develop the novel and generic machine learning model to predict entrepreneurial firms’ valuation (i.e., pre-money valuation); (2) evaluate the relative importance of each feature in the predicting model; and (3) unravel the relationship between input features of interest and entrepreneurial firm valuation and identify the interaction effects among those significant features.

The proposed machine learning model combines the differential evolution algorithms with the adaptive learning rate optimization algorithm to obtain the best parameters in an efficient way. The dataset mainly retrieved from the Thomson Reuters Eikon database and Orbis Patent database validates the proposed machine learning model. In addition, compared with the classical machine learning model (i.e., Least Squares (LS) algorithm, Ridge regression algorithm, Deep Neural Network (DNN) algorithm, and Random Forest (RF) algorithm), the proposed model performs best with regard to two performance metrics, namely, the correlation of determinant ($R^2$) and the root mean square error (RMSE).

Further, this paper attempts to increase the interpretation of the machine learning model in two ways. The partial derivate method is applied to estimate the relative importance of each feature, whereas the partial dependence plot (PDP) visualizes the specific relationships among significant features and the entrepreneurial firms’ valuation. In doing so, this paper figures out the complex relationship between firm valuation and its features and thereby provides decision-making guidance for entrepreneurial firms, VC investors, and policymakers.

The rest of this article proceeds as follows. The methodological approach is introduced in Section 2, including differential evolution algorithm, adaptive learning rate optimization algorithm, partial derivates methods, and partial dependence plot. Section 3 illustrates the data, features, model performance metrics, and baselines used in this paper. Section 4 first compares the results of the proposed model and the baseline models and then interprets the results further by performing feature contribution

---

1In the VC research field, internal rate of return, also called IRR, is often used by VC investors to evaluate the investment opportunity of programs (Gompers et al., 2020).
ranking and plotting the partial dependence between input features of interest and the output variable. Finally, in Section 5, the discussion of the conclusion, contributions, implications, and future directions are outlined.

2 | PROPOSED INTEGRATED MACHINE LEARNING MODEL

In this section, the proposed integrated machine learning model will be displayed with its merits. In particular, first, a new differential evolutionary neural network algorithm is incorporated into the neural network model to better identify the hyperparameters and network architecture. Second, the adaptive learning rate optimization algorithm is used to facilitate the training process. Third, this study employs a partial derivatives method to confirm the feature contribution ranking sorted by importance and partial dependence plot to figure out the relationships between input features and the output variable, which can provide guidance for entrepreneurial firm valuation prediction.

2.1 | Differential evolution algorithm

Machine learning especially deep neural network learning has been regarded as a very powerful way in many domains due to its strong capabilities in data analysis and pattern recognition. To get the optimized neural network, training the neural network is essential to get the best topological structure (i.e., the number of neurons in every hidden layer and the total number of hidden layers) and hyperparameters (i.e., epoch size and regularization parameter). Currently, the trial-and-error method and optimization algorithm are usually used to address the above issue since there is no exact theory to guide the determination of an optimal neural network (Suwansawat & Einstein, 2006). In particular, trial-and-error has low efficiency and has been proved to obtain suboptimal solutions. In contrast, the optimization algorithm performs better since it is able to improve convergence speed and reliability of computing significantly. The differential evolution algorithm has proved itself to outperform other optimization algorithms in training neural networks since it is easy to use, has a faster convergence speed, and has better robustness properties (Mohamed & Mohamed, 2019). In this case, the differential evolution algorithm was selected to determine the best topological structure and hyperparameters in this paper.

The differential evolution algorithm developed by Storn and Price is a heuristic and population-based global optimization method over continuous space in nature (Storn & Price, 1997). Owing to its excellent capabilities in solving numerical optimization problems, it has become prevalent in multiple fields. This technique mainly consists of four parts in the implementation process: initialization, mutation operation, crossover operation, and selection operation, which are described as follows.

2.1.1 | Initialization

The initial population was created to establish the beginning of the training process. Let $N_p$ denote the population size, $\text{rand}(0, 1)$ mean a random value between 0 and 1, and $G$ be the number of generations. According to prior research, the initial population can be described using the following formula:

$$x_{i,0} = x_{i,\text{min}} + \text{rand}(0, 1) \cdot (x_{i,\text{max}} - x_{i,\text{min}}), j = 1, 2, ..., D,$$

where $D$ denotes the number of dimensions of parameter vectors, $i$ and $j$ represent the $j$-th component of the $i$-th individual in each generation, and 0 refers to the initial generation. $x_{i,\text{min}}$ and $x_{i,\text{max}}$ are the minimum and maximum parameter bounds, respectively.

2.1.2 | Mutation

Let $v_{i,m}$ be the mutant vector and $x_{i,m}$ be the target vector. After initialization, the mutant vector is created according to the most frequently used mutation strategy (i.e., DE/rand/1 strategy).

$$v_{i,m} = x_{r1,m} + F(x_{r2,m} - x_{r3,m}), r_1 \neq r_2 \neq r_3 \neq i.$$

$r_1, r_2, r_3$ are randomly chosen indices in the range $[1, 2, ..., N_p]$, and $F \in [0, 2]$ is the scale factor that is corresponding to the amplification of the difference vector $(x_{r2,m} - x_{r3,m})$.

2.1.3 | Crossover

The binominal crossover is incorporated in this crossover process as the differential evolution algorithm adopts this scheme. The trial vector $u_{i,m}$ is obtained from the combination of the target vector and mutant vector. In this case, the trial vector $u_{i,m}$ is applied as

$$u_{i,m} = \begin{cases} v_{i,m}, & \text{if } (\text{rand}(0, 1) \leq CR) \text{ or } (j = \text{rand}) \\ x_{i,m}, & \text{otherwise} \end{cases}$$
where \( \text{rand}(0,1) \) denotes a uniformly random number between 0 and 1 and CR \((\text{CR} \in [0,1])\) is the crossover rate that indicates the degree that individual components inherit from the mutant vector. The perform of \( j = j_{\text{rand}} \) is to make sure that at least one component of the mutant vector can pass on to the trial vector.

### 2.1.4 Selection

This operation follows the greedy selection strategy. Put another way, if the fitness function of the trial vector \( u_{i,m} \) cannot be as good as the old vector \( x_{i,m} \), then \( x_{i,m} \) will be retained. Otherwise, the trial vector will replace the old one and be adopted as the next generation \( m + 1 \). The selection procedure can be described as follows.

\[
x_{i,m+1} = \begin{cases} u_{i,m}, & \text{if } f(u_{i,m}) \leq f(x_{i,m}) \\ x_{i,m}, & \text{otherwise} \end{cases}
\]  

where \( f \) represents the fitness function applied to the individual in the generation.

### 2.2 Adaptive learning rate optimization algorithm

It is well known that the machine learning model consumes lots of computing resources especially big data used for the analysis (Jordan & Mitchell, 2015). In this case, the learning rate is of great importance to the performance of the proposed model. Basically, a lower learning rate can make the training process more reliable but time-consuming. In contrast, a higher learning rate can save lots of time but usually make the training model not converge (Njock et al., 2020). Indeed, the learning rate is considered to be very hard to set among all the hyperparameters. An adaptive learning rate optimization algorithm proposed by Kingma and Ba (2014) can better address the aforementioned problem. This algorithm only requires low memory and is able to facilitate the training process at the same time. Thus, this algorithm was employed to improve the computing efficiency.

### 2.3 Partial derivate method

In order to determine the feature contribution analysis, the coefficient of every feature needs to be calculated. Several methods can be used to do this work. However, following previous research (Gevrey et al., 2003), the partial derivate method is regarded as the most appropriate method since it can generate robust results and calculate the coefficients with deep hidden layers.

The partial derivatives method is implemented in two steps: (1) Get the first-order derivatives, and (2) add all the derivatives of all the hidden layers. Following previous research (Dimopoulos et al., 1995), the calculations for the partial derivative method are given as follows.

\[
d^l_i = \frac{\partial \hat{y}}{\partial x_k} = \sum_{i=1}^{m} \frac{\partial \hat{y}}{\partial z_{l-1}^i} \frac{\partial z_{l-1}^i}{\partial x_k} = \sum_{i=1}^{m} \left( w_{kj}^l g'(z_{l-1}^i) \frac{\partial z_{l-1}^i}{\partial x_k} \right)
\]

\[
d^l_i = \frac{\partial \hat{y}}{\partial x_k} = \frac{\partial \hat{y}}{\partial x_k} = \frac{\partial \hat{y}}{\partial x_k} = g'(z_{l-1}^i) d_{l-1}^i,
\]

\[
c_k = \frac{\partial \hat{y}}{\partial x_k} = g'(z_{l-1}^i) d_{l-1}^i,
\]

\[
C_k = \sum_{p=1}^{n} \left( \sum_{j=1}^{m} \frac{\partial \hat{y}}{\partial x_k} \right)^2,
\]

where \( d^l_i \) and \( C^l \) represent the differential coefficients, \( n \) denotes the total number of observations, and \( x_k \) indicates the \( k \)-th input parameter. Finally, a greater \( C^l \) value means more importance to the feature and a bigger contribution to the prediction model.

### 2.4 Partial dependence plot

The partial dependence function proposed by Friedman (2001) can be used to open the black box of machine learning and thus increase the interpretability with regard to the model further. Accordingly, partial dependence plots (PDPs) are utilized to visualize the predicted effect of one or two features after eliminating the remaining features. In particular, one-way PDP describes the relationship between the input features of interest and the average predicted output variable, while two-way PDP depicts the interaction effect among two input features of interest. From these PDPs, it is easy to justify the exact relationship between the average predicted output variable and the input features (namely, linear, monotonic, or even more complex).

Let \( X_s \) be a feature set, \( S \subseteq \{1, 2, ..., p\} \), and let \( X_c \) denote the remaining features. That is to say, \( X_c \) is the complement subset of \( X_s \). Hence, \( S \cup C = \{1, 2, ..., p\} \) and \( S \cap C = \emptyset \). Following previous research (Molnar et al., 2021), the partial dependence function of machine learning model \( f(X) \) for regression is given as
Monte Carlo integration is employed to estimate the partial dependence function, which is shown as follows:

\[ f_S(X_S) = \frac{1}{n} \sum_{i=1}^{n} f(X_S, X_C), \quad (10) \]

where \( n \) denotes the number of observations in the training dataset. \( X_S \) and \( X_C \) together make up the complete feature space \( X \). From this formula, the average marginal effect of machine learning model \( f \) can be captured when the specific value(s) to the chosen feature set \( X_S \).

2.5 | Steps of proposed scheme implementation

The detailed implementation procedure of the proposed integrated machine learning scheme is described as follows (see Figure 1).

Step 1. Selecting all the relevant features according to prior research;

Step 2. Collecting the data and pre-processing the data (including cleaning the data, and preparing the initial input values);

Step 3. Implement the Adam-ENN model to get the hyperparameters and structures of the neural network;

Step 4. Constructing the neural network models and training the model until its objective function value or the generation reaches the set value;

Step 5. Evaluating the proposed model and baseline models with performance metrics;

Step 6. Applying the partial derivative method to do feature contribution analysis;

Step 7. Employing the partial dependence plot to depict the relationship between the input features of interest and the output variable, namely, entrepreneurial firm valuation.

3 | CASE STUDY

3.1 | Data

The VC and entrepreneurial firm data used in this paper were mainly compiled from the Thomson Reuters Eikon.
database. In particular, this database purports to offer comprehensive VC investment data on the global market from 1961 to the present and detailed information about entrepreneurial firms, including firms’ websites, found date, address, website, and funding amount. The patent data were retrieved from the Orbis patent database. The dataset was also enriched by the Qcc database, which specializes in information about Chinese entrepreneurial firms’ credit, products, co-founder/team, and so on.

The choice is to only consider China in this paper for two reasons. Previous empirical studies pay more attention to the developed market, such as the United States. However, there is little research about emerging markets even though there are some differences between emerging markets and developed markets in terms of entrepreneurship financing environment. In recent years, China has become the second-largest VC market in the world (Wang, 2017), which offers a good context and representative sample to observe the entrepreneurship financing obtained from VC investors.

The data used in this paper range from January 1, 1998, to December 31, 2018. Left censoring should not be a major concern since Chinese VC investments occurred rarely before 1998. Following prior research (Liu & Maula, 2016), firm deal was set as the unit of our sample. In so doing, the conditions and characteristics in the deal level can be easily controlled (Gu & Lu, 2014).

The initial sample has 463,240 firm-deal observations, which included 15,077 unique VC investors and 90,003 unique entrepreneurial firms between 1998 and 2018. To construct the sample, we first excluded (1) VC investors and entrepreneurial firms whose names were undisclosed; (2) observations with missing values; (3) the observations in which the variable term “Firm status = defunct”; (4) the funding amount is undisclosed or missing; and limited (5) entrepreneurial firm country = China. Moreover, to eliminate the effects of varying investment motivation and learning of syndication networks, this paper only focuses on the first-round financing (Sorenson & Stuart, 2008). Thus, the final dataset contains 912 unique VC investments made into 314 unique Chinese entrepreneurial firms in the first-round entrepreneurship financing.

3.2 | Features

Consistent with prior research (Alexy et al., 2012; Block et al., 2014; Kaplan et al., 2009), the features that were employed in training or testing models can be divided into five types: (1) VC firm characteristics; (2) VC investors’ social capital; (3) entrepreneurial firm (also called venture) characteristics; (4) funding round characteristics; and (5) founder/team characteristics. All features used in this paper are listed in Table 1.

### Table 1: List of features predicting entrepreneurial firm valuation

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>VC firm characteristics</td>
<td>VC age</td>
</tr>
<tr>
<td></td>
<td>VC size</td>
</tr>
<tr>
<td></td>
<td>VC experience</td>
</tr>
<tr>
<td></td>
<td>Investor type</td>
</tr>
<tr>
<td></td>
<td>Foreign</td>
</tr>
<tr>
<td>VC social capital</td>
<td>Centrality</td>
</tr>
<tr>
<td></td>
<td>Clustering coefficient (CC)</td>
</tr>
<tr>
<td></td>
<td>Structural hole (SH)</td>
</tr>
<tr>
<td></td>
<td>Small world quotient (SWQ)</td>
</tr>
<tr>
<td>Venture characteristics</td>
<td>Patent number</td>
</tr>
<tr>
<td></td>
<td>Business model</td>
</tr>
<tr>
<td></td>
<td>Venture stage</td>
</tr>
<tr>
<td></td>
<td>Venture age</td>
</tr>
<tr>
<td></td>
<td>Venture city</td>
</tr>
<tr>
<td></td>
<td>Venture industry</td>
</tr>
<tr>
<td>Funding round characteristics</td>
<td>Syndicate size</td>
</tr>
<tr>
<td></td>
<td>Fund inflows</td>
</tr>
<tr>
<td></td>
<td>Market size</td>
</tr>
<tr>
<td>Founder/team</td>
<td>Founder education</td>
</tr>
<tr>
<td></td>
<td>Founder financing experience (FFE)</td>
</tr>
</tbody>
</table>
the rest is assigned to 0. Given foreign and domestic VC firms use different investment strategies (Gu & Lu, 2014), a dummy variable called “foreign” was created. If the focal VC firm is located in one country outside China, 1 was assigned to this variable and 0 otherwise.

3.2.2 | VC investors’ social capital

VC investors are embedded in networks, that is, “bound by their current and past investments into webs of relationships” (Hochberg et al., 2007, p. 251). Thus, the social capital of VC investors underlying the VC network generates significant impacts on the ventures. It should be noted that one node refers to a VC investor while the tie indicates that two VC investors located at two ends of the tie co-invest with each other in a given time in the syndication network (Ter Wal et al., 2016). In this study, the following social network features are included: network position (eigenvector centrality), the clustering of VC networks (clustering coefficient), network non-redundant resources (structural holes), and small-world network metric (small world quotient).

3.2.3 | Venture characteristics

The number of patents, the proxy of ventures’ innovation capability (Guan et al., 2015; McCarthy & Aalbers, 2016), is regarded as one of the important factors to impact venture valuation (Tumasjan et al., 2021). Following prior research (Chemmanur et al., 2014; Yan et al., 2018, 2020), we construct the Patent number variable employed the patent application year since the application year is actually closer to the time of innovation of the focal firm than the grant year and thus is more suitable to indicate firm innovation. Besides, the business model of ventures is another significant consideration, which is frequently mentioned in the business plan. According to previous research (Evans & Gawer, 2016), a dummy variable was constructed and was set to one if the venture was a platform company and zero otherwise. The ventures’ stage is more likely to affect the focal VC firm’s investment decisions (Shane & Stuart, 2002); therefore, the ventures’ stage was incorporated into the prediction model, namely, the seed stage and growth stage. Moreover, ventures’ age, city, and industry were also selected as venture-related features to predict pre-money valuation (Alexy et al., 2012; Hsu & Ziedonis, 2013).

3.2.4 | Funding round characteristics

Syndicate size (measured by the number of syndicated VC investors in a financing deal) was included as it can convey the signal to attract potential VC firms (Sorensen & Stuart, 2001; Tumasjan et al., 2021) and thereby impact the valuation of the entrepreneurial firm. Fund inflows were considered because prior research has shown that fund inflows to the venture can increase the valuation (Gompers & Lerner, 2000). Market size reflects the market reaction to the venture and has been confirmed to influence VC firms to evaluate the value of the venture in previous research (Miloud et al., 2012).

3.2.5 | Founder/team characteristics

Finally, founder/team-related features are described as follows. Founder/team education was calculated as the total number of people who can get a master’s degree or PhD degree. Founder financing experience is a dummy variable that equals one if the founder had received the funding from VC investors before this funding round.

3.3 | Metrics

In order to evaluate the effectiveness of the machine learning model for the prediction of continuous variables, two common-used performance metrics were employed (Gao et al., 2020; Granata et al., 2017), that is, the correlation of determination ($R^2$) and the root mean square error (RMSE). The range of $R^2$ goes from 0 to 1. The higher value of $R^2$ indicates better model performance. On the contrary, the lower the value of RMSE, the better the model performance is. The mathematical formulas are presented as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \quad (12)$$
where \( n \) represents the total number of observations, \( y_i \) is the \( i \)th target value, and \( \tilde{y}_i \) is the \( i \)th predicted value of pre-money valuation. \( \bar{y} \) denotes the mean value of all the target values.

### 3.4 Baselines

In this paper, entrepreneurial firm valuation (pre-money) prediction can be readily mapped into the regression problem. According to previous research (Ponta et al., 2022; Shalev-Shwartz & Ben-David, 2014), the Least Squares (LS) algorithm is regarded as the most classical and simple algorithm to predict regression problems. Therefore, it was employed as the first baseline model in this paper. We chose Ridge regression as our second baseline model owing to its good performance in avoiding the overfitting issue of the least square model (Hoerl & Kennard, 1970). Compared with traditional machine learning algorithms, Deep Neural Network (DNN) has shown great promise in modeling complex and non-linear relations in the model (Njock et al., 2020). Obviously, the proposed integrated machine learning model was built based on DNN, which is regarded as the power tool to deal with complex problems. Therefore, DNN was chosen as the third baseline model. Meanwhile, considering that the RF algorithm is regarded as the state-of-the-art general-purpose algorithm to deal with regression or classification problems (Wainberg et al., 2016), the RF algorithm was chosen as the last baseline method. Scikit-learn software package was employed to implement the aforementioned baseline algorithms.

The following sections will present the comparison results of five algorithms to verify the predominance of the proposed integrated machine learning model.

### 4 Modeling Results and Analysis

In this section, the data processed in Section 3.1 and features described in Section 3.2 are used to train the model and the results of the proposed machine learning model and baseline models (i.e., LS, Ridge regression, DNN, and RF algorithm) are compared and discussed. Specifically, the training process and experimental setup are presented at first. Then, the model performance and feature contribution are analyzed. Finally, the complex relationships between the input features of interest and entrepreneurial firm valuation are disclosed by Partial Dependence Plot (PDP).

#### 4.1 Experimental setup

According to previous research (Ponta et al., 2022), all the data are randomly divided into two parts: 70% of the data are utilized to train the model, while the remaining data (30%) are utilized to test the model. In order to get robust results, all the experiments were repeated 30 times. Table 2 presents the list of parameters utilized in the proposed integrated machine learning model, which is implemented in Fortran.

Eventually, with the increase of generations, the objective function (i.e., log-cash loss function) declines sharply and finally converged at the 18th generation, which confirms the high efficiency and usefulness of the differential evolution algorithm as shown in Figure 2.

Accordingly, the topological structure of the neural network and hyperparameters of the best neural network are identified. That is, the neurons of the input layer are

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Epoch size (E)</td>
<td>[1,000, 2,500]</td>
</tr>
<tr>
<td>2 Regularization (( \lambda ))</td>
<td>[0, 100]</td>
</tr>
<tr>
<td>3 Population size (P)</td>
<td>30</td>
</tr>
<tr>
<td>4 Number of generations (NG)</td>
<td>50</td>
</tr>
<tr>
<td>5 Number of neurons/layer (Ne)</td>
<td>[5,30]</td>
</tr>
<tr>
<td>6 Number of hidden layers (Nh)</td>
<td>[1, 7]</td>
</tr>
<tr>
<td>7 Mutation rate (M)</td>
<td>0.5</td>
</tr>
<tr>
<td>8 Crossover rate (CR)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

![FIGURE 2 Objective function value with generation](https://onlinelibrary.wiley.com/doi/10.1002/for.2912)
20, 6 hidden layers, 29 neurons in every hidden layer, 1 neuron in the output layer (see Figure 3), and epoch size = 1927, regularization parameter = 0.00002.

4.2 | Model performance

Figures 4 and 5 show the prediction performance achieved by the proposed model and baseline models (i.e., LS, Ridge regression, DNN, and RF algorithm). In particular, Figure 4 presents the correlation of determination ($R^2$) of the proposed integrated machine learning model and baseline models. It can be easily observed that the proposed machine learning model significantly enhanced over the baselines. Compared with the baseline models, the correlation of determination is improved by $24.2\%$, $23.7\%$, $1.8\%$, and $3.3\%$ in the training dataset and $29.3\%$, $37.1\%$, $26.9\%$, and $26.7\%$ in the testing dataset.

As shown in Figure 5, the RMSE of the proposed integrated machine learning model is lower than the RMSE of LS, Ridge, DNN, and RF algorithms both in the training and testing phases, which further confirms the good prediction performance of our proposed model.
4.3 Feature contribution analysis

Previous research posits that machine learning algorithms work well but exist model interpretability problems, which is especially true when it comes to the neural network (Rebala et al., 2019). One important way to solve this problem is to perform the feature contribution analysis (Ponta et al., 2022). Gevrey et al. (2003) investigated seven ways to address this issue, and the results indicated that the partial derivatives method generally produced the most robust results. Hence, this paper employs this method to determine the importance of features.

Figure 6 presents the results of feature ranking sorted by importance for the first-round entrepreneurial firm valuation. As shown in Figure 6, syndicate size (i.e., the number of VC investors in the funding round) plays the most important role in predicting entrepreneurial firm valuation, which further confirms that the Chinese VC market exists the “circle” phenomenon (Gu & Liu, 2019). More syndicated VC investors convey good signals that the entrepreneurial firm behaves excellent in its industry and such endorsement effect in return helps the entrepreneurial firm obtain a high valuation in the first-round entrepreneurship financing.

Business model performs better than the founder financing experience, which is consistent with previous research (Sharchilev et al., 2018). That said, the business model of the entrepreneurial firm is more important than the team characteristic in the first-round financing, which also supplements the answers to the debate regarding whether investors should bet on the jockey or the horse (Kaplan et al., 2009). It is noteworthy that VC social capital characteristics stand out among these features. In particular, a VC investor with more structural holes can receive more novel resources and timely information, which can make a reasonable evaluation of the firm value. Clustering and eigenvector centrality of VC syndication networks imply the degree of high trust (Zhang et al., 2021) and prestige of VC investors (Hochberg et al., 2007), respectively. The presence of such influential VC investors can produce a significant impact on the valuation of entrepreneurial firms. Fund inflows have a non-negligible influence on entrepreneurial firm valuation, which is consistent with prior research (Gompers & Lerner, 2000; Tumasjan et al., 2021). Interestingly, number of patents cannot crack the top 10 features, which demonstrates that the initial phase of the entrepreneurial firm seldomly exhibits strong innovation capability, especially in the first-round entrepreneurship financing.

4.4 Visualizing the relationship between the outcome variable and features (partial dependence plot)

In order to open up the black box of machine learning model further, partial dependence plot was employed to unveil the relationship between important input features of interest and entrepreneurial firm valuation and the interaction effect between any two features. All the features displayed in this section are shown on a normalization scale between 0 and 1.

As shown in Figure 7, syndicate size, business model, founder financing experience, and patent number are positively related to the valuation of entrepreneurial firms. In particular, syndicate size increases steadily at first, and after 0.8, grows sharply due to the syndicated benefits such as low trust cost and information sharing. Since business model is a dummy variable, the dependence displayed in Figure 7 just confirms that high-valuation entrepreneurial firms are usually platform-based. It is easy to understand that the founder who had VC financing experience before is more attractive to VC investors and thus can help the venture obtain a high valuation. Although number of patents (i.e., the proxy of innovation capacity) cannot play the pioneering role in the first-round financing, stronger innovation performance still attracts better and higher firm value in the valuation phase.

Fund inflows and VC investors’ social network characteristics demonstrate the nonlinear and complex relationships with an entrepreneurial firm valuation, which is explicitly displayed in Figure 7. Specifically speaking, as fund inflows increase, the valuation reaches a certain low level, then keeps constant beyond that point (approximately 0.17), and increases steadily at last. This finding reminds the danger of oversimplifying the relationship between fund inflows to VC investors and
FIGURE 7 One-factor partial dependence plot of entrepreneurial firm valuation on input features of interest. The pikes on the x axis denote the fractile of input feature values and represent the data density.
entrepreneurial firm valuation. The social network features (viz., structural holes, clustering coefficient, and eigenvector centrality) of syndication networks behave a non-linear (non-monotonic) impact on the entrepreneurial firm valuation, which confirms the complex impacts of the social capital of VC investors on entrepreneurial firm valuation further.

From Figure 8a, the partial dependence is high, while the entrepreneurial firm is a platform company (business model) and becomes older (venture age). In Figure 8, the patent number represents the capability of the entrepreneurial firm, while founder financing experience indicates the ability of the founder/team of this entrepreneurial firm to some degree. As exhibited in Figure 8b, the founder experience can hardly interact with patents number to improve the entrepreneurial firm valuation due to the non-obvious patent applications at the beginning of entrepreneurship financing. A steady increase is followed by a slight decline in the later phase and eventually reaches the highest partial dependence (Figure 8c). Put another way, syndicate size has a positive interaction effect with VC experience on the entrepreneurial firm’s valuation. The last picture (Figure 8d) demonstrates the mutual impacts among syndication network features, and the results confirm the combined effects. Specifically speaking, the partial dependence is high when the structural holes range from 0.4 to 0.8, and the clustering coefficient is larger than 0.35 (or smaller than 0.10). Overall, these two-factor PDP analyses provide intriguing results for the entrepreneurship financing research field.

5 CONCLUSION

This paper proposes an integrated machine learning scheme so as to predict the valuation of entrepreneurial firms, which are conducted by VC investors. We incorporate the differential evolution algorithm into the traditional deep neural network model and further employ the adaptive learning rate optimization algorithm to facilitate the training process. Compared with the baseline models (Least Squares (LS) algorithm, Ridge regression algorithm, DNN, and RF algorithm), the proposed
machine learning model could achieve better convergence, which shows the higher efficiency of the model. No over-fitting issue exists during the learning process due to the applied regularization approach. The good generalization ability of the proposed machine learning model was confirmed by a centralized relative percentage error. Furthermore, the prediction performance of the training process is 98.30%, while the value in the testing process is 92.40%. Feature contribution analysis indicates that the first key feature responsible for entrepreneurial firm valuation is syndicate size, which typically reflects the importance of signal conveyed from existing VC investors. VC investors' social capitals are very important determinants to predict entrepreneurial firm value. Interestingly, innovation capability seems not so important in early-stage entrepreneurial firm valuation since entrepreneurial firms in nascent stages lack the invention records and thus are not able to convey the useful firm quality signal.

This paper contributes to venture capital and entrepreneurial finance literature in several ways. First, this paper develops a valuation model for early-stage VC investors, which is applied to entrepreneurial firms in the first-round financing. Extant research mainly focuses on the impact of one feature/a few features on entrepreneurial firm financing (Alexy et al., 2012; Block et al., 2014; Tumasjan et al., 2021; Zheng et al., 2010). Few studies compare all the features and identify the feature contribution of them. As yet, only simple linear regressions or limited features have been considered to investigate the relationship between entrepreneurial financing and its features. With the emerging state of machine learning techniques (Liu et al., 2022; Qiu et al., 2021), this study develops a novel machine learning model to predict entrepreneurial firm valuation and thereby provides a more holistic viewpoint to entrepreneurial financing literature. Second, this study adds to the venture capital literature by indicating and clarifying the relative importance of the business model (Horse) and founder/team (Jockey). Our finding indicates that the business model of entrepreneurial firms behaves more important than the founder/team of entrepreneurial firms especially in the first-round financing when VC investors evaluate the value of entrepreneurial firms, which offers a new perspective for entrepreneurs to pay more attention to the business (product/service) than the “relationship” and “friendship” (Gu & Liu, 2019). In so doing, this finding also allows us to add to the ongoing debate in relation to the importance of business model and founder/team on firm evaluation (Kaplan et al., 2009; Sharchilev et al., 2018). Third, we advance entrepreneurial finance research by introducing and investigating the role of VC investors’ social capital in firm valuation. In particular, this research not merely reminds the danger of oversimplifying the relationship between the social capital underlying syndication network and entrepreneurial firm valuation. Importantly, we also provide some evidence that different kinds of social capital exist complex interaction effects on firm valuation. Put another way, clustering (clustering coefficient) and brokerage advantage (structural holes) do not always have a positive interaction effect on firm valuation. Hence, this study contributes to the emerging research stream, which explicitly considers the non-negligible impact of VC investors’ social capital on entrepreneurial financing (Milosevic et al., 2020; Zhang et al., 2021).

This study makes practical implications for VC investors, entrepreneurs, and policymakers. For VC investors, this paper provides an insightful way to make an investment decision. They should pay more attention to the business model of entrepreneurial firms, especially in the first-round investment. On the contrary, the innovation capability of entrepreneurial firms seems not so crucial in the initial phase of investment. For entrepreneurs, the findings of this paper can help them to identify the important determinants to attain high firm value from VC investors. The elite cliques/investment circles can bring seductive social capitals but those advantages generate non-linear impacts on entrepreneurial firm valuation. For policymakers, this paper can assure the existence and importance of the investment circle and “guanxi” in the context of the Chinese VC investment environment. They should take action to control this phenomenon and help entrepreneurial firms that do not have the opportunity to access VC to receive funding from VC investors to sustain their development.

While offering multiple insights into entrepreneurial financing research, this article has some limitations that open up promising avenues for future research. First, in this paper, only Chinese entrepreneurial firms are considered since China has become the second-largest VC market in the world so as to become one of the representative countries to conduct entrepreneurship research. Future research can extend this work to other countries, such as America, European countries, and other Asian countries. Second, owing to the limitation of available data, only the social capital of VC investors is considered. However, it is possible that entrepreneurs are able to build up some partnership networks (i.e., alliance networks) among themselves. It would be another interesting future inquiry to explore whether and how such social capital of entrepreneurs may influence entrepreneurial firm valuation for VC investors.

ACKNOWLEDGMENTS
The authors would like to thank the editor Prof. Derek Bunn, ADM Sowmya Murali, and two anonymous...
referees for their constructive and valuable comments, which helped to improve the quality of this paper. The financial support of this article was granted by China Scholarship Council (No. 201806630043). We also acknowledge the valuable and detailed feedback from Igors Skute at the University of Twente, as well as other participants at R&D Management conference 2021.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author.

ORCID
Ruling Zhang https://orcid.org/0000-0002-0423-1178

REFERENCES
syndication? Evidence from the renewable energy industry. 
Sustainability, 13(7), 4050. https://doi.org/10.3390/su13074050
Zheng, Y., Liu, J., & George, G. (2010). The dynamic impact of 
innovative capability and inter-firm network on firm valuation: 
A longitudinal study of biotechnology start-ups. Journal of Busi-
jbusvent.2009.02.001

**AUTHOR BIOGRAPHIES**

**Ruling Zhang** is a PhD student at Donghua University (Glorious Sun School of Business and Management), and a visiting PhD student at University of Groningen (Faculty of Economics and Business). Her research interests include entrepreneurship financing, venture capital, and machine learning.

**Zengrui Tian** is a professor at Glorious Sun School of Business and Management at Donghua University. She received her PhD degree from Xi’an Jiaotong University. Her research interests include entrepreneurship financing, venture capital, and firm innovation.

**Killian J. McCarthy** is an associate professor at University of Groningen (Faculty of Economics and Business). He received his PhD degree in economics from University of Groningen. His research interests include mergers and acquisitions, entrepreneurship financing, and technology management.

**Xiao Wang** currently serves as an associate professor in Innovation and Entrepreneurship at Xi’an Jiaotong-Liverpool University. He obtained his PhDs from University of Groningen and University of Chinese Academy of Sciences. His research interests include entrepreneurial ecosystem, entrepreneurship financing, and innovation.

**Kun Zhang** obtained his PhD degree from Shanghai Jiaotong University and master’s degree from Wuhan University. His current research interests include machine learning, energy policy, and civil engineering.