IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 20, 2020, accepted April 29, 2020, date of publication May 6, 2020, date of current version May 20, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2992869*

# Predicting Students' Performance With School and Family Tutoring Using Generative Adversarial Network-Based Deep Support Vector Machine

**KWOK TAI CHUI[1], (Member, IEEE), RYAN WEN LIU[2], (Member, IEEE), MINGBO ZHAO[3], (Senior Member, IEEE), AND PATRICIA ORDÓÑEZ DE PABLOS[4]**

[1]School of Science and Technology, The Open University of Hong Kong, Hong Kong
[2]Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan 430063, China
[3]School of Information Science and Technology, Donghua University, Shanghai 200051, China
[4]Department of Business Administration and Accountability, Faculty of Economics, The University of Oviedo, 33003 Oviedo, Spain

Corresponding authors: Kwok Tai Chui (jktchui@ouhk.edu.hk) and Ryan Wen Liu (wenliu@whut.edu.cn)

**ABSTRACT** It has been witnessed that supportive learning has played a crucial role in educational quality enhancement. School and family tutoring offer personalized help and provide positive feedback on students' learning. Predicting students' performance is of much interest which reflects their understanding on the subjects. Particularly it is desired students to manage well in fundamental knowledge in order to build a strong foundation for post-secondary studies and career. In this paper, improved conditional generative adversarial network based deep support vector machine (ICGAN-DSVM) algorithm has been proposed to predict students' performance under supportive learning via school and family tutoring. Owning to the nature of the students' academic dataset is generally low sample size. ICGAN-DSVM offers dual benefits for the nature of low sample size in students' academic dataset in which ICGAN increases the data volume whereas DSVM enhances the prediction accuracy with deep learning architecture. Results with 10-fold cross-validation show that the proposed ICGAN-DSVM yields specificity, sensitivity and area under the receiver operating characteristic curve (AUC) of 0.968, 0.971 and 0.954 respectively. Results also suggest that incorporating both school and family tutoring into the prediction model could further improve the performance compared with only school tutoring and only family tutoring. To show the necessity of ICGAN and DSVM, comparison has been made between ICGAN and traditional conditional generative adversarial network (CGAN). Also, the proposed kernel design via heuristic based multiple kernel learning (MKL) is compared with typical kernels including linear, radial basis function (RBF), polynomial and sigmoid. The prediction of student's performance with and without GAN is presented which is followed by comparison with DSVM and with traditional SVM. The proposed ICGAN-DSVM outperforms related works by 8-29% in terms of performance indicators specificity, sensitivity and AUC.

**INDEX TERMS** Generative adversarial network, students' academic performance, deep support vector machine, supportive learning.

## I. INTRODUCTION

Learning analytics [1], [2] and supportive learning [3], [4] have become emerging research areas in today's era of big data and artificial intelligence to facilitate students' learning. Student education is vital to the sustainable development of society because students learn knowledge and abilities to contribute the community. There are many students who have progressed to higher level or graduate every year. However, some students marginally pass the course and some fail from the course are usually required to have a compulsory retake. Many research works have detailed the analysis of the interrelated negative effects on students who have marginally passed or failed the course. These can be explained in three perspectives. Students may experience the reduction of confidence [5] and even suffer from depression [6] attributed to dissatisfactory course grade. The deferral and early school leaving (or termination) of students' studies may increase the

The associate editor coordinating the review of this manuscript and approving it for publication was Miguel Jesus Torres Ruiz.

workload of staff and expenditure [7]. In addition, not only the reputation of the school [8] but also the social capital [9] will be lowered as a result of students receiving fail grade.

Traditional learning between teachers and students in normal class and supportive learning accomplish one another in recent decades. The advanced development of information and communications technology (ICT) architecture and technologies offers plenty of opportunities via e-learning [10] and virtual reality education [11]. In this paper, the focus will be on another supportive learning environment: shadow education via school tutoring and family tutoring. Its prevalence has become a worldwide phenomenon [12], [13]. Students often attend after-class school tutoring, could be small group or individual. Family members may also provide support in tutoring. These ways devote efforts in building closer relationship with learners which help in fine-tuning and customizing the best approach for learners. Particularly, how learners learn are very important so that proper guidance can be given.

Predicting students' performance is desired so that proper follow-up actions could be setup to help students who are in-need. In literature, various machine learning algorithms have been proposed and evaluated using real-world datasets. Researchers have analyzed students' heterogeneity for feature extraction [14]. Prediction models were implemented using four common machine learning algorithms, JRip, sequential minimal optimization, C4.5 and Naïve-Bayes. All algorithms have achieved similar prediction accuracy of 80%. Another work in [15] proposed a gradient boosting machine algorithm to predict students' performance at the end of the academic year. The attributes age, school, neighborhood, absence and grade were found to be effective measures in students' performance. Results tested by two datasets were 86% and 89% in accuracy. However, the positive and negative classes were significantly unbalanced, with ratio of 1:7. Attention was drawn into the feature extraction process, 42 features belonging to one of the feature groups grades, status, load, family background as well as course difficulty, level, performance and specification, were analyzed [16]. A preliminary study of prediction algorithm for students' performance was carried out using various methods, random, forest, decision tree, support vector machine (SVM) and gradient boosting. area under the receiver operating characteristic curve (AUC) is between 0.5 and 0.877 under different testing datasets and approaches. In [17], support vector machine, neural network and decision were applied to predict students' performance associated with daily internet usage. Support vector machine achieved the highest average accuracy among three, which is about 70%. Differed from shallow learning in [14]–[17], deep learning approach based on deep artificial neural network was employed [18]. Results indicated that this deep learning approach outperformed support vector machine and logistic regression by 4.3% and 8.6% respectively. Here are the recommended state-of-the-art articles [19], [20] for readers who are interested in the overview of algorithms for students' performance prediction.

Existing works [14]–[18] possessed a common idea of analyzing the optimal feature vector from the dataset. Taking the review articles [19], [20] into account, to the best of our knowledge, there has no consideration on the prediction of students' performance under shadow education environment, that is school tutoring and family tutoring. On the other hand, the machine learning algorithms were mainly shallow learning approach because there is usually small data volume in education datasets. Also, in general, there is room for improvement in the prediction accuracy. A recent work [18] using deep learning was suggested an improvement of accuracy by 4.3% compared to support vector machine with traditional kernel function. The improvement may become insignificant if customized kernel or multiple kernel learning approach is adopted.

To address the limitations. This paper has proposed an improved conditional generative adversarial network based deep support vector machine (ICGAN-DSVM) algorithm. ICGAN aims at addressing the issue of low data volume by mimicking new training dataset whereas DSVM extends SVM from shallow learning to deep learning. DSVM takes the advantage in small dataset, as a key difference comparing with traditional deep neural network.

The contributions of this paper are summarized as (i) School tutoring and family tutoring have been taken into consideration in the formulation of prediction students' performance, which is first of its kind; (ii) ICGAN has demonstrated its effectiveness in generating new training data compared with traditional CGAN which facilitates new research direction in learning analytics; (iii) DSVM is employed which takes the advantage in small-sized educational data environment.; and (iv) the proposed ICGAN-DSVM algorithm improves the specificity, sensitivity and AUC by about 8-29% comparing with existing works.

The rest of the paper is organized as follows. Section II presents the dataset and section III illustrates the methodology of proposed ICGAN-DSVM. Thorough analysis on the effectiveness of ICGAN and DSVM as well as comparison to existing methods will be given in Section IV. At last, conclusion is drawn in Section V.

## II. STUDENT PERFORMANCE DATASET

The dataset for student performance prediction was retrieved from [21]. It is comprised of two classes from 788 students (i) Portuguese language class of 649 records; and (ii) Mathematics class of 395 records. The dataset has 33 attributes in which 9 of them are related to school tutoring and family tutoring. The attributes are parent's cohabitation status, mother's education, mother's job, father's education, father's job, student's guardian, quality of family relationships, school educational support and family educational support. The rest, 29 of the attributes were collected by questionnaire and the remaining were from school reports. These attributes are student's school, student's sex, student's age, student's home address type, family size, reason to choose this school, home to school travel time, weekly study time, number of past class failures, extra paid classes within the course subject, extracurricular activities, attended nursery school, wants to take higher education, internet access at home, with a romantic relationship, free time after school, going out with friends, workday alcohol consumption, weekend alcohol consumption, current health status, number of school absences, first period grade, second period grade and final grade.

To investigate the influence of school tutoring and family tutoring, three scenarios will be considered. Scenario 1: consider only school tutoring; Scenario 2: consider only family tutoring; Scenario 3: consider both school tutoring and family tutoring.

## III. METHODOLOGY OF ICGAN-DSVM

In this section, the methodology of ICGAN-DSVM will be discussed. First, the rationale and the details of ICGAN are presented as the method to generate more training student performance data. It is followed by DSVM which is responsible for the prediction model of students' performance.

### A. GENERATE NEW TRAINING DATA WITH ICGAN

Aforesaid, GAN is chosen to increase the data volume of the dataset. The generator and discriminator compete to achieve the Nash equilibrium in the training stage. It is typical to have small-sized educational dataset in practice. Generally, continuous data collection at the very beginning, that is when learners were young is difficult to achieve. A recent review article [22] has summarized the recent progress and various approaches of GAN. There are four categories named convolution-based, conditional-based, autoencoder-based and objective function optimization-based methods.

In this paper, we adopt conditional-based GAN (CGAN). In the original GAN, the random noise vector (as the generator's input) is unimpeded which may cause fatal theory corruption. To address this limitation, conditional variable is introduced in the generator and discriminator. In literature, there are three highly cited (over 1000 citations from Google Scholar) approaches for CGAN, the original form CGAN [23], InfoGAN [24] and auxiliary classifier GAN (ACGAN) [25]. Fig. 1 shows the conceptual flows of existing approaches for better illustration. Denote the symbols n as the noise vector, a as conditional variable, G as generator, X as data distribution, D as discriminator, Q as additional network. These approaches have well been demonstrated effectively in various applications. G captures the data distribution whereas D estimates the probability that a sample came from the training data rather than G. Both G and D are conditioned. D could determine whether the data is from G or original dataset. Generated data has certain bias but is acceptable if it is low. One idea is to introduce a constraint to maximize the diversity because diversity and bias are inversely correlated. We have confirmed the prediction model via ICGAN has low bias in generated data by examining the density of the data. Therefore, the introduction of constraint of diversity is avoided.

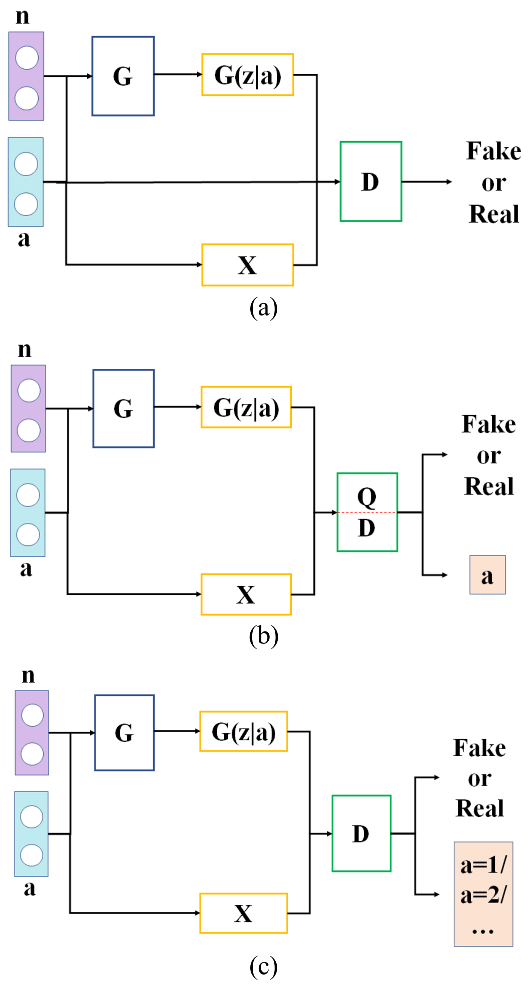



**FIGURE 1.** Conceptual flow of existing CGAN approaches. (a) CGAN. (b) InfoGAN. (c) ACGAN.

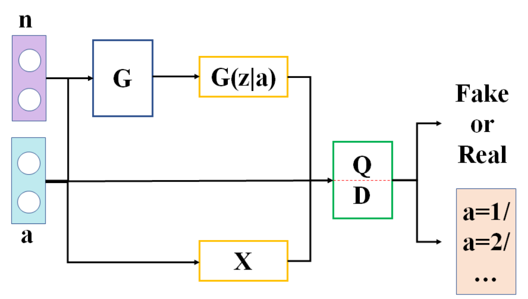



**FIGURE 2.** Conceptual flow of proposed ICGAN.

Each of existing works [23]–[25] has its own advantages that leads to superior performance. As a result, we have proposed an improved CGAN (ICGAN) that combines the existing architectures of CGAN, InfoGAN and ACGAN. Fig. 2 presents the conceptual flow of ICGAN. ICGAN incorporates the ideas of (i) introducing conditional variable a to discriminator; (ii) adding additional network Q along with discriminator; and (iii) assigning label to every generated

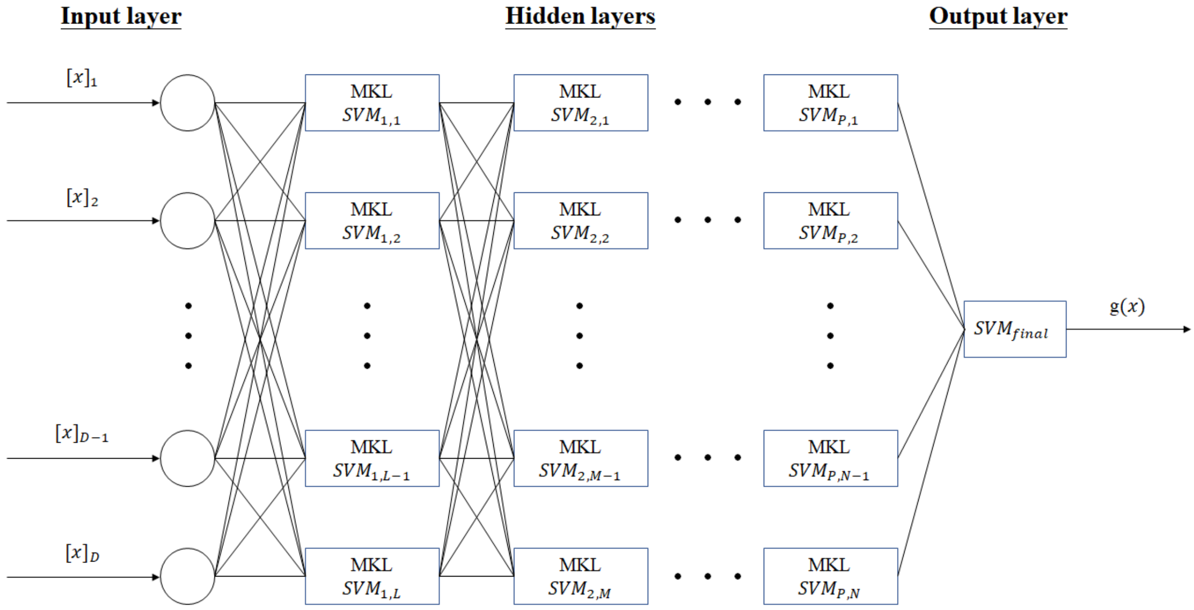



**FIGURE 3. Conceptual flow of DSVM using MKL.**

sample. The objective function consists of three parts:

$$L_{source} = E\left[logP\left(S = real|X_{real}\right)\right] + E\left[logP\left(S = fake|X_{fake}\right)\right] \quad (1)$$

$$L_{class} = E\left[logP\left(C = a|X_{real}\right)\right] + E\left[logP\left(C = a|X_{fake}\right)\right] \quad (2)$$

$$I\left(a, G(n,a)\right) = E_{x\sim G(n,a)}[E_{a\sim P(a|x)}\left[\log \mathrm{Q(a|x)}\right]] + \mathrm{H(a)} \quad (3)$$

The formulation is intended to maximize $L_{source}+L_{class}-\lambda$ I(a,G(n,a)) for discriminator and maximize $L_{class} - L_{source} - \lambda$ I(a,G(n,a)) for generator. $\lambda$ is the hyperparameter and I(a,G(n,a)) is the mutual information between a and G(n,a).

### B. STUDENT PERFORMANCE PREDICTION MODEL WITH DSVM

The prediction model for student performance is implemented using DSVM architecture. In general, it has multiple hidden layers of SVM and an output layer of SVM. Compared to other deep learning architectures like deep neural network, DSVM takes several advantages like (i) able to manage problem of very large input vectors and small-sized training dataset; (ii) the design of kernel functions is more flexible; and (iii) the output layer SVM has strong regularization power to avoid over-fitting.

The flow of DSVM is shown in Fig. 3. Denote some integers D, L, M, N and P. It is worth mentioning that the number of hidden layers is arbitrary. In Section IV, analysis will be carried out on the selection of number of layers by grid search. Authors have suggested to use grid search to reduce the computational power for optimal search. Small number of hidden layers is normally obtained in real-world applications, further increase of hidden layers may deteriorate the performance of the model.

Typical kernel functions adopted in SVM include linear, radial basis function (RBF), $p^{th}$ order polynomial, and sigmoid kernels. The major research concern raised by researchers is these kernels could not yield optimal performance in all applications. Therefore, customizing kernel to every application is desired, multiple kernel learning (MKL) has received much of attention [26]–[28].

In this paper, the DSVM utilizes MKL to combine typical kernel functions. The combination of kernel functions to form resultant kernel function must obey Mercer's theorem [29]. The classifier can achieve better performance by taking the advantages from each kernel. To align with the major focus of related works, authors consider linear, RBF, $p^{th}$ order polynomial and sigmoid kernels. They are defined with (4)-(7) respectively using the notation of kernel function $K(x_1,x_2)$ with inner product $\langle x_1,x_2 \rangle$.

$$K_1\left(x_1, x_2\right) = \langle x_1, x_2 \rangle \quad (4)$$

$$K_2\left(x_1, x_2\right) = exp\left(||x_1 - x_2||^2/2\sigma\right) \quad (5)$$

$$K_3\left(x_1, x_2\right) = \left(x_1, x_2 + c\right)^p \quad (6)$$

$$K_4\left(x_1, x_2\right) = tanh\left(x_1, x_2 + c\right) \quad (7)$$

where $\sigma$ and c are real numbers and p is positive integer.

Heuristic approach is adopted for MKL. The basic formulations are summarized as follows [30]. Define the kernel alignment $F(K_i,q)$ between kernel matrix $K_i$ and label set z.

$$F\left(K_i, q\right) = \langle K_i, qq^T \rangle_F / \sqrt{\langle K_i K_i \rangle_F \langle qq^T, qq^T \rangle_F} \quad (8)$$

Trivially, if $K_i$ has a large alignment to z, there is a large contribution on resultant kernel. Therefore, the F-heuristic is defined as:

$$\mu_i = F(K_i, q)/\sum\nolimits_{i=1}^{4} F(K_i, q) \quad (9)$$

It can be further incorporated with the consideration of mean square error (MSE). F-heuristic becomes M-heuristic.

$$\mu_i = \sum_{i=1}^{4} (M_i - M_j) / \sum_{j=1}^{4} \sum_{i=1}^{4} (M_i - M_j) \quad (10)$$

In every SVM as in Fig. 3, the designed kernel by MKL may differ from each other as an extension to existing heuristic approach.

It is worth noting that the proposed algorithm ICGAN-DSVM is comprised of two parts. The complexity of ICGAN is comparable to existing CGAN, InfoGAN and ACGAN because ICGAN is the combination of these ideas. When it comes to DSVM, each SVM follows the complexity of $O(n^2 p + n^3)$ and $O(n_{sv} p)$ in training and prediction stage, where n is the number of samples, p is the number of features and $n_{sv}$ is the number of support vectors. Since DSVM takes the advantages in small size problems, the requirement of computational power is much less than that of typical deep learning algorithms, like convolutional neural network.

## IV. ANALYSIS AND RESULTS OF ICGAN-DSVM

The analysis of the effectiveness of proposed ICGAN-DSVM will be discussed in four parts: (i) The performance of the proposed ICGAN-DSVM is evaluated with school tutoring and/or family tutoring; (ii) Compare the performance between ICGAN and typical CGAN approaches; (iii) Compare the performance between kernel using heuristic based MKL and typical kernel functions; and (iv) Compare the performance between proposed ICGAN-DSVM and related works.

### A. FORMANCE EVALUATION OF ICGAN-DSVM

Grid search method has been chosen to select the number of hidden layers in the DSVM architecture. The range of hidden layers is from 1 to 6. Consideration will be made between ICGAN-DSVM and DSVM on the benefit of newly generated data by ICGAN. Also, three scenarios are setup: Scenario 1: consider only school tutoring; Scenario 2: consider only family tutoring; Scenario 3: consider both school tutoring and family tutoring.

Table 1 summarizes the specificity, sensitivity and AUC of DSVM and ICGAN-DSVM with varying number of hidden layers under Scenario 1. Specificity and sensitivity are defined as follows.

$$Specificity = TN/N_n \quad (11)$$

$$Sensitivity = TP/N_p \quad (12)$$

where TN is true negative, $N_n$ is number of negative samples, TP is true positive and $N_p$ is number of positive samples. AUC is the area under the 1-Specificity and Sensitivity curve.

K-fold cross-validation with has been adopted which K = 10 is a good choice supported by various related works [31]–[33]. Similarly, the performance of DSVM and ICGAN-DSVM in Scenario 2 and Scenario 3 is presented in Table 2 and Table 3 respectively.

**TABLE 1. Scenario 1: Performance of DSVM versus ICGAN-DSVM.**

| Number of hidden layers | Specificity | Sensitivity | AUC |
|---|---|---|---|
| | (DSVM / ICGAN-DSVM) | | |
| 1 | 0.900 / 0.916 | 0.901 / 0.914 | 0.890 / 0.903 |
| 2 | 0.912 / 0.923 | 0.916 / 0.928 | 0.899 / 0.910 |
| 3 | 0.926 / 0.943 | 0.924 / 0.945 | 0.912 / 0.926 |
| 4 | 0.921 / 0.929 | 0.917 / 0.934 | 0.905 / 0.916 |
| 5 | 0.916 / 0.926 | 0.919 / 0.932 | 0.903 / 0.913 |
| 6 | 0.913 / 0.924 | 0.915 / 0.925 | 0.900 / 0.909 |

**TABLE 2. Scenario 2: Performance of DSVM versus ICGAN-DSVM.**

| Number of hidden layers | Specificity | Sensitivity | AUC |
|---|---|---|---|
| | (DSVM / ICGAN-DSVM) | | |
| 1 | 0.886 / 0.902 | 0.890 / 0.904 | 0.875 / 0.888 |
| 2 | 0.893 / 0.913 | 0.898 / 0.910 | 0.884 / 0.897 |
| 3 | 0.912 / 0.926 | 0.911 / 0.921 | 0.900 / 0.908 |
| 4 | 0.904 / 0.921 | 0.907 / 0.916 | 0.892 / 0.905 |
| 5 | 0.901 / 0.915 | 0.896 / 0.913 | 0.886 / 0.903 |
| 6 | 0.897 / 0.913 | 0.900 / 0.915 | 0.885 / 0.902 |

**TABLE 3. Scenario 3: Performance of DSVM versus ICGAN-DSVM.**

| Number of hidden layers | Specificity | Sensitivity | AUC |
|---|---|---|---|
| | (DSVM / ICGAN-DSVM) | | |
| 1 | 0.908 / 0.934 | 0.906 / 0.939 | 0.895 / 0.921 |
| 2 | 0.921 / 0.943 | 0.922 / 0.946 | 0.908 / 0.928 |
| 3 | 0.933 / 0.968 | 0.931 / 0.971 | 0.915 / 0.954 |
| 4 | 0.929 / 0.959 | 0.928 / 0.956 | 0.913 / 0.944 |
| 5 | 0.923 / 0.952 | 0.924 / 0.950 | 0.912 / 0.936 |
| 6 | 0.918 / 0.947 | 0.916 / 0.944 | 0.903 / 0.929 |

The observations of the results in Table 1-III are summarized as follows.

(i) Averaging the results, ICGAN-DSVM improves (compared with DSVM) the specificity, sensitivity and AUC by (1.33, 1.56, 1.26)% in Scenario 1. Similarly, the improvements are (1.80, 1.43, 1.52)% in Scenario 2 and (3.09, 3.29, 3.05)% in Scenario 3.

(ii) Best performance in terms of specificity, sensitivity and AUC can be obtained with three hidden layers in all scenarios. Further increase of the number of hidden layers decrease the performance. The best performance of proposed ICGAN-DSVM yields specificity of 0.968, sensitivity of 0.971 and AUC of 0.954.

(iii) The prediction model works the best in Scenario 3, which is followed by Scenario 1 and Scenario 2 respectively. The reasons could be explained by the fact that both school tutoring and family tutoring help improving students' learning and thus the prediction model should include both these factors. Compared Scenario 1 and Scenario 2, the suggestion is school tutoring is slightly more beneficial compared to family tutoring. This could be explained by school tutors have more experience due to their daily job nature.

### B. COMPARISON BETWEEN ICGAN AND EXISTING CGANs

To study the effectiveness of proposed ICGAN, it is compared with traditional CGAN [23], InfoGAN [24] and ACGAN [25]. The comparison is shown in Fig. 4. Likewise,

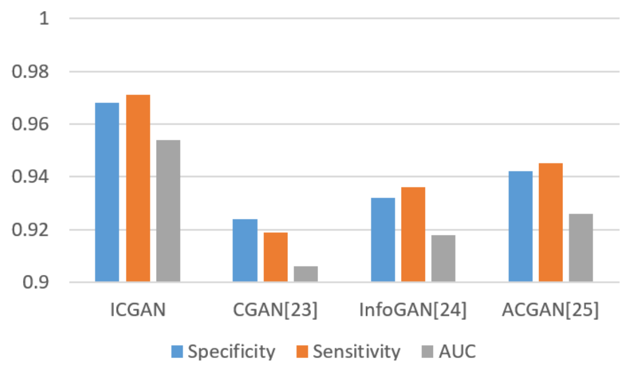



**FIGURE 4.** Performance comparison between ICGAN and existing CGAN approaches.

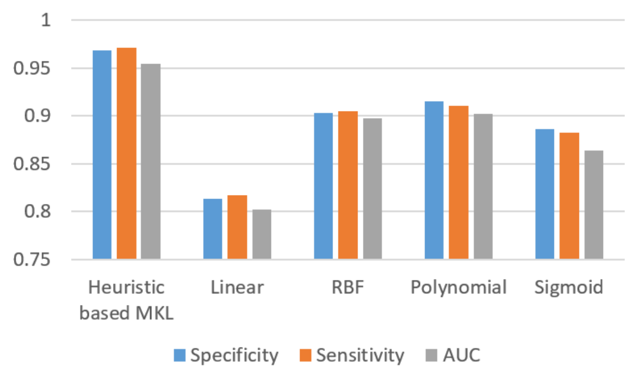



**FIGURE 5.** Performance comparison between heuristic based MKL and typical kernel functions.

the performance indicators are specificity, sensitivity and AUC, as of the averaged results of 10-fold cross-validation. It can be seen that the proposed ICGAN achieves highest specificity, sensitivity and AUC. The percentage improvement of specificity is 2.76-4.76%, 2.75-5.66%, 3.02-5.30% in specificity, sensitivity and AUC respectively. It shows that the merging of existing approaches could improve the accuracy of prediction model by taking advantages from each approach.

### C. COMPARISON BETWEEN HEURISTIC BASED MKL AND TYPICAL KERNEL FUNCTIONS

Evaluation is moved to the heuristic based MKL. It is compared with typical kernel functions, that are stand-alone linear, RBF, polynomial and sigmoid kernel functions. Fig. 5 shows the results of heuristic based MKL versus typical kernel functions. Results revealed that heuristic based MKL obtains highest specificity, sensitivity and AUC compared to existing kernels. The improvement is 5.79-19.1%, 6.70-18.8%, and 5.76-19.0% in terms of specificity, sensitivity and AUC. It shows that combining kernels can take advantages from each of the kernel to improve the prediction performance.

### D. COMPARISON BETWEEN ICGAN-DSVM AND RELATED WORKS

The last part of the analysis is performance comparison between ICGAN-DSVM and related works [14]–[18] which

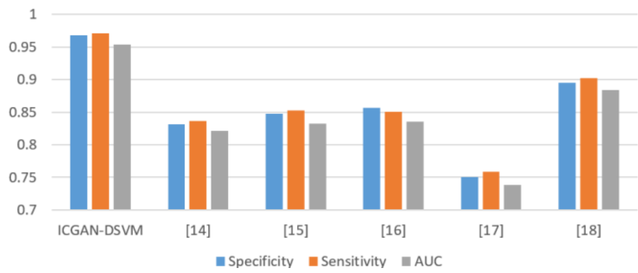



**FIGURE 6.** Performance comparison between ICGAN-DSVM and related works.

results have been summarized in Fig. 6. Results indicate that the proposed ICGAN-DSVM has best performance. The percentage improvement is 8.16-29.0%, 7.65-27.9%, and 7.92-29.3% in specificity, sensitivity and AUC respectively. Authors have suggested the following reasons for better performance of proposed work (i) Shallow learning [14]–[17] may achieve lower accuracy because it may not learn some of the hidden characteristics from the data; (ii) The deep artificial neural network in [18] is traditional deep learning technique that are basically with large-sized dataset [34]–[36] and may not suit well to the nature of low data volume application; (iii) The proposed ICGAN effectively generates new samples whereas DSVM takes the advantages in low data volume environment. Given the customized kernel has been designed based on heuristic based MKL, the proposed algorithm achieves better performance in terms of specificity, sensitivity and AUC.

## V. CONCLUSION

In this paper, authors have proposed an ICGAN-DSVM algorithm to improve the prediction accuracy of students' performance. Results have revealed its effectiveness by comparing between ICGAN and existing CGANs, between heuristic based MKL and typical kernel functions as well as between ICGAN-DSVM and related works.

Authors anticipate that current research will provide insights to programme leaders, teachers, tutors and family member when making decisions concerning supportive learning in education. The prediction of at-risk students could benefit students who are in need, thus increasing their success rate of passing the course and avoiding passing with a marginal grade. In addition, it is suggested to consider the introduction of GAN when it comes to small-sized machine learning problems. The generation of new data will benefit the implementation of model.

Future research directions are suggested as follows. The proposed method can be further applied to other educational and learning analytics datasets to demonstrate the benefit of ICGAN in generating extra data for training and DSVM is preferred to address small-sized machine learning problems compared to deep neural network. In addition, if ICGAN can be enhanced in a way that it can generate much more data without scarifying the model performance. This allows the formulation of advanced early students' performance prediction model that can estimate the performance of students multiple times per semester.

## REFERENCES


[1] A. Moubayed, M. Injadat, A. Bou Nassif, H. Lutfiyya, and A. Shami, ''E-learning: Challenges and research opportunities using machine learning & data analytics,'' *IEEE Access*, vol. 6, pp. 39117–39138, 2018.

[2] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, ''Using learning analytics to scale the provision of personalised feedback,'' *Brit. J. Educ. Technol.*, vol. 50, no. 1, pp. 128–138, Jan. 2019.

[3] C. Batanero, L. de-Marcos, J. Holvikivi, J. R. Hilera, and S. Oton, ''Effects of new supportive technologies for blind and deaf engineering students in online learning,'' *IEEE Trans. Educ.*, vol. 62, no. 4, pp. 270–277, Nov. 2019.

[4] C. Wang, H.-C.-K. Hsu, E. M. Bonem, J. D. Moss, S. Yu, D. B. Nelson, and C. Levesque-Bristol, ''Need satisfaction and need dissatisfaction: A comparative study of online and face-to-face learning contexts,'' *Comput. Hum. Behav.*, vol. 95, pp. 114–125, Jun. 2019.

[5] K. A. Zapko, M. L. G. Ferranto, R. Blasiman, and D. Shelestak, ''Evaluating best educational practices, student satisfaction, and self-confidence in simulation: A descriptive study,'' *Nurse Edu. Today*, vol. 60, pp. 28–34, Jan. 2018.

[6] C. Fiorilli, S. De Stasio, C. Di Chiacchio, A. Pepe, and K. Salmela-Aro, ''School burnout, depressive symptoms and engagement: Their combined effect on student achievement,'' *Int. J. Educ. Res.*, vol. 84, pp. 1–12, Jan. 2017.

[7] M. S. M. Momo, S. J. Cabus, K. De Witte, and W. Groot, ''A systematic review of the literature on the causes of early school leaving in Africa and Asia,'' *Rev. Edu.*, vol. 7, no. 3, pp. 496–522, Oct. 2019.

[8] V. M. López-Pastor, P. Pintor, B. Muros, and G. Webb, ''Formative assessment strategies and their effect on student performance and on student and tutor workload: The results of research projects undertaken in preparation for greater convergence of universities in Spain within the European higher education area (EHEA),'' *J. Further Higher Edu.*, vol. 37, no. 2, pp. 163–180, Mar. 2013.

[9] D. Noyens, V. Donche, L. Coertjens, T. van Daal, and P. Van Petegem, ''The directional links between students' academic motivation and social integration during the first year of higher education,'' *Eur. J. Psychol. Edu.*, vol. 34, no. 1, pp. 67–86, Jan. 2019.

[10] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani, and N. Alalwan, ''E-learning continuance satisfaction in higher education: A unified perspective from instructors and students,'' *Stud. High Educ.*, vol. 43, no. 11, pp. 2003–2019, Nov. 2018.

[11] G. Makransky and L. Lilleholt, ''A structural equation modeling investigation of the emotional value of immersive virtual reality in education,'' *Educ. Technol. Res. Develop.*, vol. 66, no. 5, pp. 1141–1164, Oct. 2018.

[12] Y. C. Kim and J.-H. Jung, ''Conceptualizing shadow curriculum: Definition, features and the changing landscapes of learning cultures,'' *J. Curriculum Stud.*, vol. 51, no. 2, pp. 141–161, Mar. 2019.

[13] C. Doherty and K. Dooley, ''Responsibilising parents: The nudge towards shadow tutoring,'' *Brit. J. Sociology Edu.*, vol. 39, no. 4, pp. 551–566, May 2018.

[14] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, and Q. Long, ''Predicting academic performance by considering student heterogeneity,'' *Knowl.-Based Syst.*, vol. 161, pp. 134–146, Dec. 2018.

[15] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. V. Erven, ''Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,'' *J. Bus. Res.*, vol. 94, pp. 335–343, Jan. 2019.

[16] A. Polyzou and G. Karypis, ''Feature extraction for next-term prediction of poor student performance,'' *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 237–248, Apr./Jun. 2019.

[17] X. Xu, J. Wang, H. Peng, and R. Wu, ''Prediction of academic performance associated with Internet usage behaviors using machine learning algorithms,'' *Comput. Hum. Behav.*, vol. 98, pp. 166–173, Sep. 2019.

[18] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, ''Predicting academic performance of students from VLE big data using deep learning models,'' *Comput. Hum. Behav.*, vol. 104, Mar. 2020, Art. no. 106189.

[19] N. Tomasevic, N. Gvozdenovic, and S. Vranes, ''An overview and comparison of supervised data mining techniques for student exam performance prediction,'' *Comput. Edu.*, vol. 143, Jan. 2020, Art. no. 103676.

[20] F. Widyahastuti and V. U. Tjhin, ''Performance prediction in online discussion forum: State-of-the-art and comparative analysis,'' *Procedia Comput. Sci.*, vol. 135, pp. 302–314, Jan. 2018.

[21] P. Cortez and A. Silva, ''Using data mining to predict secondary school student performance,'' in *Proc. FUBUTEC*, Porto, Portugal, 2008, pp. 5–12.

[22] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, ''Recent progress on generative adversarial networks (GANs): A survey,'' *IEEE Access*, vol. 7, pp. 36322–36333, Mar. 2019.

[23] M. Mirza and S. Osindero, ''Conditional generative adversarial nets,'' Nov. 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784

[24] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, ''Infogan: Interpretable representation learning by information maximizing generative adversarial nets,'' in *Proc. NIPS*, Barcelona, Spain, 2016, pp. 2172–2180.

[25] A. Odena, C. Olah, and J. Shlens, ''Conditional image synthesis with auxiliary classifier GANs,'' in *Proc. ICML*, Sydney, NSW, Australia 2017, pp. 2642–2651.

[26] M. Gönen and E. Alpaydın, ''Multiple kernel learning algorithms,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011.

[27] C. Liu, L. Tang, and J. Liu, ''Least squares support vector machine with self-organizing multiple kernel learning and sparsity,'' *Neurocomputing*, vol. 331, pp. 493–504, Feb. 2019.

[28] Z. Wang, Z. Zhu, and D. Li, ''Collaborative and geometric multi-kernel learning for multi-class classification,'' *Pattern Recognit.*, vol. 99, Mar. 2020, Art. no. 107050.

[29] R. Herbrich, *Learning Kernel Classifiers Theory and Algorithms*. London, U.K.: MIT Press, 2002.

[30] S. Qiu and T. Lane, ''A framework for multiple kernel support vector regression and its applications to siRNA efficacy prediction,'' *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 6, no. 2, pp. 190–199, Apr. 2009.

[31] Y. Kokkinos and K. G. Margaritis, ''Managing the computational cost of model selection and cross-validation in extreme learning machines via Cholesky, SVD, QR and Eigen decompositions,'' *Neurocomputing*, vol. 295, pp. 29–45, Jun. 2018.

[32] J. Lei, ''Cross-validation with confidence,'' *J. Amer. Stat. Assoc.*, vol. 114, pp. 1–20, Oct. 2019.

[33] V. Wottschel, D. T. Chard, C. Enzinger, M. Filippi, J. L. Frederiksen, C. Gasperini, A. Giorgio, M. A. Rocca, A. Rovira, N. De Stefano, M. Tintoré, D. C. Alexander, F. Barkhof, and O. Ciccarelli, ''SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis,'' *NeuroImage, Clin.*, vol. 24, Jan. 2019, Art. no. 102011.

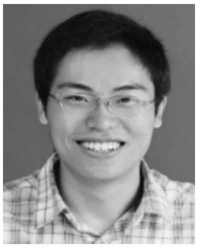
[34] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, ''A survey on deep learning for big data,'' *Inf. Fusion*, vol. 42, pp. 146–157, Jul. 2018.

[35] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, ''Object detection with deep learning: A review,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[36] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, ''Urban big data fusion based on deep learning: An overview,'' *Inf. Fusion*, vol. 53, pp. 123–133, Jan. 2020.


**KWOK TAI CHUI** (Member, IEEE) received the B.Eng. degree in electronic and communication engineering–business intelligence minor and the Ph.D. degree from the City University of Hong Kong.

He had industry experience as a Senior Data Scientist with the Internet of Things (IoT) Company. He joined the Department of Technology, School of Science and Technology, The Open University of Hong Kong, as a Research Assistant Professor. He has more than 45 research publications, including edited books, book chapters, journal articles, and conference papers. His research interests include computational intelligence, data science, energy monitoring and management, intelligent transportation, smart metering, healthcare, machine learning algorithms, and optimization. He has served as various editorial positions in ESCI/SCIE-listed journals, including a Managing Editor for the *International Journal on Semantic Web and Information Systems*, a Topic Editor for *Sensors*, and an Associate Editor for the *International Journal of Energy Optimization and Engineering*.



**RYAN WEN LIU** (Member, IEEE) received the B.Sc. degree (Hons.) in information and computing science from the Department of Mathematics, Wuhan University of Technology, Wuhan, China, in 2009, and the Ph.D. degree in mathematical imaging from The Chinese University of Hong Kong, Hong Kong, in 2015. He was a Visiting Professor with the Agency for Science, Technology and Research, Singapore. He is currently an Associate Professor with the School of Navigation, Wuhan University of Technology, and a Visiting Scholar with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include mathematical imaging, computer vision, trajectory data mining, and computational navigation sciences. He is an Associate Editor of the *International Journal on Semantic Web and Information Systems*.

**MINGBO ZHAO** (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, in January 2013.

He was with the City University of Hong Kong as a Postdoctoral Researcher. He is currently a Full Professor with Donghua University, Shanghai, China. He has authored or coauthored over 50 technical articles published at prestigious international journals and conference, including the IEEE Transactions on Knowledge and Data Engineering, the IEEE Transactions on Image Processing, *ACM Transactions on Intelligent Systems and Technology*, the IEEE Transactions on Industrial Informatics, the IEEE Transactions on Industrial Electronics, *Pattern Recognition*, and *Neural Networks*. His current research interests include pattern recognition and machine learning.

**PATRICIA ORDÓÑEZ DE PABLOS** is a Professor with the Department of Business Administration and Accountability, Faculty of Economics, The University of Oviedo, Spain. Her teaching and research interests focus on the areas of strategic management, knowledge management, and intellectual capital. She serves as an Associate Editor for *Behaviour and Information Technology*. She is the Editor-in-Chief of the *International Journal of Learning and Intellectual Capital* (IJLIC), the *International Journal of Asian Business and Information Management* (IJABIM), and the *Journal of Science and Technology Policy Management* (JSTPM), as well as an Editor for a number of IGI Global book publications and full book series.