IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 18, 2018, accepted September 11, 2018, date of publication September 26, 2018, date of current version October 17, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870877

A Novel Technique for the Evaluation of Posterior Probabilities of Student Cognitive Skills

SADIQUE AHMAD[1], KAN LI[1], ADNAN AMIN[2], AND SALABAT KHAN[1]

[1]School of Computer, Beijing Institute of Technology, Beijing 100081, China

[2]Center for Excellence in Information Technology, Institute of Management Sciences, Peshawar 25000, Pakistan

Corresponding author: Kan Li (likan@bit.edu.cn)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4172054 and in part by the National High Technology Research and Development Program of China (973 Program) under Grant 2013CB329605.

ABSTRACT To achieve excellent marks in job interviews and written examinations, a student must acquire impressive cognitive skills (CS) value. Nevertheless, the effects of frustration and CS related human factors (CSRFs) profoundly influence the student's skills during the aforementioned cognitive tasks. The recent methods present significant student's skills measurement techniques that compute the relationship among frustration, CS, and CSRF. Meanwhile, these methods become insufficient if the student's characteristics are not correctly quantized and simulated. No prior work can measure the posterior probabilities of student's CS during interviews and written examinations. In the current attempt, a novel CS measurement technique is proposed that simulates the nonlinear relationship among CS, frustration, and CSRF. First, the range of CS (0 to 20) is quantized and split into 21 periodic discrete outcomes. Proposing such range and then breaking it into components ensure the accuracy of CS prediction technique. Second, frustration is divided into four effects that have a strong association with CS. Third, the latent variable CSRF is split into two factors (mother job and exposure). Frustration and CSRF are referred to as umbrellas, while the effects of frustration and the factors of CSRF are referred to as layers of the umbrellas. The technique estimated the posterior probabilities of CS outcomes under the umbrella of frustration effects. Furthermore, the obtained posterior probabilities of CS are refined under the umbrella of CSRF factors. During the extensive experiment, the proposed technique is tested on two datasets. The obtained results show that the relationship among CS, frustration, and CSRF is successfully simulated because we achieved significant prediction accuracy. In the end, we compared the proposed approach with the prior competitive methods which concluded this study.

INDEX TERMS Cognitive skills prediction, posterior probabilities of cognitive skills, cognitive skills measurement, student's skills measurement.

I. INTRODUCTION

Measuring a student's time-varying Cognitive Skill (CS) (knowledge state) is essential to pinpoint the student's strengths and deficiencies during job interviews and written examination. CS follows study material as well as student's family-related characteristics during the aforementioned critical tasks. These particular characteristics strongly influence the performance and academic achievements of an individual. Besides, by CS prediction techniques, a student can acquire experience about his or her weak performance (e.g., after interview or examination). The primary objective of the current study is to develop an algorithmic method that can track the particular CS value of a student. Psychology, neuroscience, biological science, and cognitive science have hundreds of studies on the relationship between CS and other human factors (experience, education, exposure, and emotions, etc.) of a student. These methods are insignificant to achieve a unified CS measurement technique that efficiently tracks the real-time knowledge state of a student.

We have investigated a recent technique which used a student's performance dataset for the prediction of student's skills [1]. This method predicted the student's skills by using the results of the students' solved exercises (e.g., exercise solutions can be correct or incorrect). According to this method, the CS of a student increases by correctly answering the exercises. The skill of the student has no specific limit because it increases monotonically with a correct exercise solution. This method presented an important technique to discover the CS of a student. Besides, this study has some limitations because it has not considered those student's characteristics that can affect CS during interviews or written examinations. Therefore, the recent method has no explicit solution which can correctly quantize CS and then calculate its posterior probabilities under the umbrellas of student's characteristics. These student's characteristics are referred to as CS Related Factors (CSRF) and frustration effects. Furthermore, there is no formal definition of the range of student's CS which can ensure the accuracy of CS measurement. With no specific range of CS, we cannot categorize the student's skills to different levels as excellent, good, or low CS. During interviews and written examinations, frustration is a significant factor that negatively effects the particular CS of an individual [2]. The severe conflicting effects of frustration profoundly influence the values of student's CS, e.g., long drive, traveling or long time waiting for interview or examination can increase the severity of the adverse effects of frustration. Thus, the algorithmic solution for tracking the affected value of the CS has a significant role during interviews or written examinations. Secondly, CSRF as exposure and mother profession can also increase or decrease CS of a student [3]. Many other human factors can affect CS, but the primary goal of the current study is to provide an algorithmic method for the simulation of the nonlinear relationship between CS, frustration, and CSRF.

Ahmad *et al.* [4] proposed a novel technique for the simulation of the nonlinear association between CS and basic human factors (aging, infection, emotions, awareness, personality, education, and experience). This study has some limitations to achieve the goal of skills prediction during interviews and written examinations, i.e., (1) possibilities of errors during the simulation of the large set of human factors, (2) complications and possibilities of errors in the novel equations development. Thus, these recent methods are insufficient to address the following challenges.

- Accurately quantize the CS and student's characteristics (frustration effects and CSRF).
- Predict the skills of students by simulating the nonlinear relationship between CS and student's characteristics.

The current attempt has initiated the computation of student's skills by evaluating the correlation between CS, CSRF, and frustration effects. It particularly estimates the posterior probabilities of student's CS. Solving these particular challenges can increase the accuracy of CS measurement. As a first contribution, the method has initiated the quantization of CS. It has proposed a unique range for CS (0 to 20) and then split it into 21 periodic discrete outcomes (with a period of 1). This range of quantization has increased the accuracy of the CS measurement because it has enabled us to propose a prediction technique that calculates component-wise (CS periodic outcomes) posterior probabilities of student's CS. This particular range has some other advantages for student's skills prediction approach, e.g., the posterior probability of each component of CS is more maintainable and testable. As a second contribution, frustration is divided into four effects (aggression, giving up, loss of self-confidence and stress). Each effect of frustration has performed a particular action on posterior probabilities of the CS outcomes. The classification of frustration is achieved to increase the performance accuracy of the proposed approach.

Thirdly, the CSRF is divided into two variables (mother job and exposure of the individual). Mother job has four outcomes, i.e., (1) services (e.g., administrative or police), (2) teacher, (3) health (healthcare related), and (4) at home. The exposure has two types of outcomes, i.e., urban and rural. Frustration effects and CSRF are working as two umbrellas under which the posterior probabilities of CS outcomes need to be estimated. Also, the particular outcomes of the CSRF variables and frustration effects are referred to as layers of these specific umbrellas. Therefore, in each layer of the proposed technique, the Bayesian Inference method is used to evaluate the posterior probabilities of CS outcomes. The probabilities computation process of the current approach is twofold. Firstly, the technique evaluates the posterior probabilities of every CS outcome with respect to each effect of frustration. Secondly, the obtained posterior probabilities are used as a set of prior probabilities in the next iterations (under the umbrella of CSRF). Thus, the technique examines the posterior probabilities of CS under each layer of the CSRF.

During the extensive experiment, the proposed technique was tested on two datasets (1: public dataset for CSRF and 2: psychological experiment based dataset for frustration effects). The obtained surprising results show that the proposed technique successfully simulated the correlation between CS, frustration, and CSRF. Also, it achieved a significant CS measurement performance accuracy. The proposed computation of posterior probabilities ensured prediction accuracy and flexibility of the current technique.

In addition, section II presents the related work. The proposed method explained in section III followed by result and discussion in section IV, and section V. The paper is concluded in section VI.

II. RELATED WORK

CS measurement needs a significant knowledge base because it overlaps between machine learning, deep learning and psychology. The method requires substantial psychological findings (that based on psychological experiments) to modulate the relationship between CS and other human factors. These findings relate CS with other human factors [5], [6]. The recent work proposed significant methods that focus on the measurement of student's CS [7], [8]. Along with the significant contributions, these methods have some deficiencies which compromise accuracy during CS measurement of a student (e.g., during job interviews or attempting written examinations). The recent methods have no correct solution that based on posterior probability evaluation technique. Lindsey *et al.* [1] proposed a method to discover the skills of a student by using students' performance data. According to this method, the expected accuracy of a student monotonically increases with correctly attempting an exercise (or a section of exercise). Here, the technique discovered skills, but it became insufficient in predicting CS during interviews and written examination because it has not defined such range that can ensure prediction accuracy. Thus, proposing a specific range for student's CS and further break it into periodic intervals plays a key role in CS measurement approaches. Ahmad *et al.* [4] presented a CS measurement technique that simulates the statistical correlation between CS and Basic Human Factors (BHF). This method has defined a unique range for CS (0 to 10). Besides, it has also proposed novel domains and ranges for BHF (each factor has different domain and range). Firstly, it estimated BHF and then calculated the values of CS using estimated values of BHF. This method cannot be used for predicting CS during job interviews and written examinations due to the three main problems. Firstly, it has a lack of in-depth quantization of CS outcomes and BHF. Secondly, it has not considered those attributes of students which continuously influence CS during the interviews and written examinations. Thirdly, it is based on the nonlinear least square method; therefore, this method is more data dependent. The nonlinear least square method based technique can compromise the accuracy due to three types of errors, i.e., mistakes in the dependent and independent variables calculations, possibilities of errors during parameters estimations, (3) prediction errors in the final mathematical model [9], [10].

As we discussed earlier that CS measurement approach overlapped between different areas of research; therefore, we obtain prior probabilities of CS outcomes from psychological studies (those studies which have psychological experiments based findings). These studies illustrated the relationship between CS and student's characteristics. The first characteristic of a student is frustration that has different effects on student's skills. A student can be frustrated due to the annoying long period of traveling or waiting for job interviews [11]. Such frustration negatively affects the skills of a student during cognitive tasks. Furthermore, the literature has shown different approaches for iterative prediction of CS, but the problem statement (measuring CS during interview and examination) made it insufficient for modulating time-varying knowledge state of a student. CS measurement needs an algorithm that iteratively refines the knowledge state of a student under the influence of their characteristics [12]. Such an algorithm can be embedded for CS evaluation in the proposed technique. We have summarized different prediction techniques [13]–[15] to embed them in the proposed technique. The problem description of measurement of posterior probabilities convinced us to use Bayesian inference in our approach. This particular method calculates the posterior probabilities of CS outcomes on different nodes of the proposed architecture. Bayesian inference is a simple technique through which we can achieve a precise solution for the iterative evaluation of posterior probabilities in a different node of the technique [16].

Selecting a posterior probability calculation method is not enough for CS prediction because we need to quantize those factors that influence a student's skills during interviews and examinations. The quantization of student's characteristics (frustration, study-related, and family-related attributes) are essential for modulating and predicting CS. Furthermore, the literature has shown that frustration effects (aggression, giving up, loss of self-confidence and stress) can successfully influence (negatively) the CS level of a student during cognitive task [17], [18]. Frustration can inhibit a student's educational (during written examinations) and job (during job interviews) opportunities [19]. In addition, the unknown connections between CS and the effects of frustration are challenging. The related study [20], [21] considered a data filling approach (an unknown relationship between factors) for the measurement of the unknown relationship between soft sets. They focused on the reliabilities of the parameters. This existing study is ineffective to measure the fundamental skills of the student because the method needs prior probabilities estimation to initiate the process of the posterior probability measurement.

Khajah *et al.* [22] demonstrated a deep learning approach for predicting student's performance. They attempted to analyze Deep Knowledge Tracing (DKT) and Bayesian Knowledge Tracing (BKT). This method is insufficient to achieve the goal of CS measurement during critical cognitive tasks (as discussed earlier). To achieve a novel technique for CS prediction need to measure and modulate the relationship between CS, frustration effects and CSRF because these latent factors play a crucial role in influencing CS. The connection between CS and frustration effects is just like a social network because their network contains physical relationship and virtual factors (cognitive levels and ideologies) [23], e.g., There is a positive effect of parents jobs (services, teachers, etc.) on student's CS [24] while negative effects of failing an interview (frustration). Therefore, the attributes that need to be quantized is frustration and its effects. As we discussed earlier that the research findings of literature related CS and frustration; therefore, we split frustration into different effects during cognitive tasks (interviewing or written examination) [2]. A student with frustration is affected by the negative impacts of the particular frustration effects. Therefore, the recent methods are insufficient to modulate the posterior probabilities of CS with respect to frustration effects. We do not have a CS measurement technique that can refine the posterior probabilities of CS in each sub-iteration of a method which shows another technical deficiency of the recent approaches.

Furthermore, the related methods are ineffective in quantizing the CSRF of a student. Talanov *et al.* [25] presented emotional state simulation model. This method simulated the relationship between fear and CS. During interviews and written examinations, the roles of frustration effects and CSRF are very sensitive as compared to fear. Frustration and fear are both negative emotions for CS, but they have different effects on the values of CS of the students. Samsonovich *et al.* [26] proposed a cognitive architecture called emotional Biologically Inspired Cognitive Architecture (eBICA). The eBICA scheme is based on three major building blocks of a cognitive model: (1) moral scheme,

(2) emotional state, and (3) emotional appraisal. This model has insufficient features (of cognitive skills measurement technique) that can evaluate the cognitive functions and processes of an individual. However, we can find different cognitive architectures that have correct solutions in the form of cognitive processes and computational techniques. The literature also has supervised and unsupervised methods that compute CS during critical circumstances [27]. The study [28], [29] illustrated the basic deficiencies in cognitive architectures. They analyzed and compared different cognitive architectures.

During the literature analysis, the evaluation processes of the current technique are threefold; 1) measuring and validating the relationship between student's characteristics and CS (this section is based on psychological studies), 2) assessing the recent methods on quantization and prediction of CS and student's characteristics, and 3) investigating the recent studies on the frustration recognition and posterior probabilities estimations. The division of frustration (into different effects) is essential in CS measurement [30]. We have hundreds of literature studies on the recognition of different emotions [31]; however, the basic need is to accurately quantize these emotions that can work as the layers of the umbrella. Moreover, the literature shows that the student's skills are not only negatively affected by emotions but also via the CSRF. Therefore, we need an iterative method that can evaluate the posterior probabilities of CS under the influence of frustration and CSRF. We explored the critical problem statement beyond the boundaries of computer sciences and machine learning because CS measurement algorithm needs the background knowledge for the prior probabilities of CS and student's characteristics. Now, we can add other study-related features to the proposed method. The existing studies manifested different collaborative filtering algorithm (matrix factorization technique) [32], but such methods are ineffective due to the problem statement of the CS measurement system. Therefore, firstly, we quantize the relationship between CS and the student's characteristics and then modulate the relationship between these particular factors.

Zhang *et al.* [33] demonstrated a multi-label metamorphic prediction approach that based on neural network. This work motivated the proposed technique to split CS, frustration, and CSRF into multiple components. Predicting CS through methods as neural network and Gauss-Newton Method (GNM) incorporate different problems. These problems are raised due to missing attributes of the data (during data collection) [34]. Therefore, such problems can lead us to various errors (issues in independent and dependent variables as well as errors in the novel equation). The proposed technique is designed to evaluate the posterior probabilities of CS outcomes that overcame the limitations of GNM and other related methods [35], [36]. The discussed previous methods have significant contributions in the field of student's skills predictions; nevertheless, these innovations are insufficient to measure and quantize the relationship between CS, CSRF and frustration effects. Moreover, these methods are also insignificant to improve the accuracy and precision of the CS prediction during cognitive tasks. Consequently, the related literature did not address the aforementioned challenges of the current approach which ultimately motivated us to develop a novel method for student's CS measurement.

III. METHOD

The primary goal of the current attempt is to simulate the nonlinear statistical association between CS and other human factors. The psychological, neuroscience and cognitive studies have illustrated that CS has a strong correlation with human factors as aging, gender, education, exposure, emotions, etc. [4], [37], [38]. We can find thousands of literature that explored the statistical connection between CS and human factors as well as how an individual CS level is affected (positively and negatively). We do not have enough algorithmic techniques to provide a reliable and error-free platform for CS measurement. Therefore, in the current attempt, we have proposed a novel approach to simulate the relationship between CS, frustration and CS Related Human Factors (CSRF). The first challenge of the current attempt is to define a specific range for student's skills because we need to achieve a significant set of CS (component-wise) probability distribution. It also ensures prediction accuracy. Thus, during quantization, we define a range (0 to 20) for CS that is further split into 21 periodic discrete outcomes (with a period of 1). Secondly, the proposed technique divides frustration into four effects, i.e., aggression, giving up, loss of self-confidence and stress. The previous studies show that the selected frustration effects have an adverse relationship with CS of an individual [39]–[41].

We need to analyze CS outcomes under the umbrella of frustration effects (effects are referred to as layers of the umbrella). Thirdly, the proposed technique selects two factors (mother job and exposure) as a CSRF. The research studies also manifest that CSRF is negatively as well as positively associated with CS. The proposed technique divides CSRF into six observable variables to simulate the relationship between CS and CSRF. The mother job has four observable variables, i.e., service, teacher, health and at home while exposure has two observable variables, i.e., urban and rural. CSRF is referred to as a second umbrella while its variables are the layers of the umbrella. Further explorations of these factors are beyond the scope of the current study. To modulate CSRF, frustration, and CS periodic outcomes, the proposed technique divides this phase into the following sections.

- Find the posterior probabilities of CS outcomes with respect to the layers of the frustration umbrella.
- Refine the obtained posterior probabilities of CS under the layers of the CSRF umbrella.

Here, in each iteration of the proposed technique, we obtain the posterior probabilities of 21 outcomes of CS. Thus, the method will choose the CS outcome of the highest posterior probability as compared to the rest of the 20 CS outcome (posterior probabilities). As discussed earlier that frustration effects and the factors of CSRF are the layers of the umbrellas; therefore the proposed method uses the Bayesian inference method to calculate the posterior probability (of each outcome of CS) under the profound influence of these layers. Firstly, the technique estimates prior probabilities of CS outcomes, and then it calculates the conditional, joint and posterior probabilities of CS outcomes. Thus, to initiate the posterior probabilities calculations, the following equation shows the prior probabilities of CS outcomes.

$$\alpha_i = P(cs_i) \tag{1}$$

$$1 - \alpha_i = P(cs_i^c) \tag{2}$$

Eq.(1) and (2) represent two events which are mutually exclusive. As we discussed earlier that CS has split into 21 periodic outcomes; therefore we have 21 iterations to calculate the posterior probability of a student. In each iteration, Eq. (1) represents the prior probability of the *ith* (α_i) outcome of CS. Eq. (2) shows the prior probability of the *jth* mutually exclusive CS outcome that is represented by $(1-\alpha_i)$. Besides, α_i represents the prior probabilities of CS outcomes ($i = 1$ to 21) while $1 - \alpha_i$ represents the prior probability of the mutually exclusive CS outcome. These both events illustrate different CS values, and its probabilities have a range from 0 to 1. As a prior, we set the initial probabilities of CS outcomes with respect to frustration and CSRF. The next step is to calculate the new conditional and joint probabilities with respect to frustration effects. To find conditional and joint probabilities of CS, we need to estimate the prior of each effect of frustration. The following equation gives the prior probability of the particular frustration effect.

$$\beta_k = P(eff_l) \tag{3}$$

Eq.(3) represents the prior probability of frustration effect. We have four effects of frustration that are represented by eff_l (where $l = 1$ to 4) while the assigned prior probability of the selected effect has shown by β_k. In each iteration, the prior probabilities of these effects are refined with respect to CS outcome because each CS outcome has a different probability. The next step is to find the conditional probabilities of each CS outcome with respect to frustration effects. The following equations give the conditional probabilities of CS outcomes.

$$eff_{con_{mn}} = P(eff_l|cs_i) = \frac{P(cs_i \bigcap eff_l)}{\alpha_i} \tag{4}$$

$$eff_{con_{mn}}^c = P(eff_l|cs_i^c) = \frac{P(cs_i^c \bigcap eff_l)}{1 - \alpha_i} \tag{5}$$

Eq. (4) and (5), represent the conditional probabilities of CS with respect to frustration effects. Eq. (4) calculates the conditional probability of CS outcome (α_i of Eq. (1)) with respect to eff_l (where $l = 1$ to 4). Eq. (5) assesses the conditional probability of the mutually exclusive CS outcome ($1 - \alpha_i$ of Eq. (2)) with respect to eff_l. Moreover, in each *nth* iteration of frustration umbrella, we have *mth* iterations for the 21 outcomes of CS. Furthermore, the joint probabilities of eff_l and CS are achieved by the following equations.

$$eff_{jo_{qr}} = P(eff_l, cs_i) = (\alpha_i) \times (eff_{con_{mn}}) \tag{6}$$

$$eff_{jo_{qr}}^c = P(eff_l, cs_i^c) = (1 - \alpha_i) \times (eff_{con_{mn}}^c) \tag{7}$$

Eq. (6) and (7) illustrate joint probabilities ($eff_{jo_{qr}}$, $eff_{jo_{qr}}^c$) of *qth* effects of *rth* CS outcome. More specifically, Eq. (6) calculates the joint probability of the set of prior probabilities which is obtained from Eq. (1). Besides, the Eq. (7) evaluates the joint probability of the set of prior probability that is collected from Eq. (2). Finally, the following equation manifests the particular posterior probability of the *ith* outcome of CS (with respect to the *lth* effect of frustration).

$$eff_{post_{xy}} = P(cs_i|eff_l) = \frac{eff_{jo_{qr}}}{eff_{jo_{qr}} + eff_{jo_{qr}}^c} \tag{8}$$

The basic purpose of this process is, to formulate different effects of student's frustration during a job interview and written examination. The proposed technique iteratively analyzes these adverse effects of frustration. The posterior probabilities of CS outcomes are re-estimated under the impact of the negative effects of frustration. Through Eq. (8), the posterior probabilities of 21 outcomes of CS are processed in different iterations and sub-iterations. It manifests that the proposed technique investigates CS with respect to the function of multiple frustration effects. During this process, we have *xth* iterations while in every iteration, the posterior probabilities ($eff_{post_{xy}}$) of CS outcomes are analyzed with respect to eff_l.

Now, the technique calculates the posterior probabilities of CS outcomes under the umbrella of CSRF. In this section, the proposed technique uses the updated posterior probabilities (which are obtained with respect to the effects of frustration) of CS outcomes as a set of prior probabilities. Therefore, the first task is to find the conditional probabilities of CS outcome with respect to CSRF's factors. As we discussed earlier that the CSRF is divided into six layers; therefore, the current process has six iterations. In each layer, the conditional probabilities of CS outcomes are estimated. This particular event is given by the following equation.

$$CSRF_{con_{uv}} = P(CSRF_h|cs_i) = \frac{P(cs_i \bigcap CSRF_h)}{\alpha_i} \tag{9}$$

$$CSRF_{con_{uv}}^c = P(CSRF_h|cs_i^c) = \frac{P(cs_i^c \bigcap CSRF_h)}{1 - \alpha_i} \tag{10}$$

Eq. (9) and (10) estimate the conditional probability of the *hth* layer of CSRF. Furthermore, Eq. (9) calculates the conditional probability of CS outcome (α_i) with respect to *hth* ($CSRF_h$) factor of CSRF while Eq. (10) calculates the conditional probability of the mutually exclusive event (given in Eq. (2)) of CS outcome. On the other hand, in each *uth* layer, we have *vth* iteration for CS outcome. Nextly, the current technique achieves the joint probabilities of CS outcomes. These particular posterior probabilities are calculated according to the frustration effects and the factors of CSRF.

$$CSRF_{jo_{cd}} = P(CSRF_h, cs_i) = (\alpha_i) \times (CSRF_{con_{uv}}) \tag{11}$$

$$CSRF_{jo_{cd}}^c = P(CSRF_h, cs_i^c) = (1 - P\alpha_i) \times (CSRF_{con_{uv}}^c) \tag{12}$$

Thus, Eq.(11) and (12) represent the joint probabilities of CSRF and CS outcomes ($CSRF_{jo_{cd}}$, $CSRF_{jo_{cd}}^c$).

Moreover, c ($c = 1$ to n) represents the iterations of CS outcomes while d ($d = 1$ to m) shows the iterations of CSRF layers. Eq. (11) calculates the joint probability of the particular CS outcomes (given by Eq. (1), i.e., α_i) while the Eq. (12) processes the joint probability of the mutually exclusive event (which is given by Eq. (2), i.e., $1 - \alpha_i$). Therefore, the posterior probabilities of CS outcomes (with respect to CSRF layers) are given by the following equation.

$$CSRF_{post_{st}} = P(cs_i|CSRF_h) = \frac{CSRF_{jo_{cd}}}{CSRF_{jo_{cd}} + CSRF^{c}_{jo_{cd}}} \quad (13)$$

Eq. (13) calculates the final posterior probabilities of CS outcomes ($CSRF_{post_{st}}$) under the umbrella of CSRF. Moreover, $CSRF_{jo_{cd}}$ represents the joint probability (given by Eq. (11)) while $CSRF^{c}_{jo_{cd}}$ represents the joint probability of the mutually exclusive event (given by Eq. (12)). This section also has two types of iterations. The s represents CS outcomes while t manifests layers of the CSRF's factors. Furthermore, the proposed technique achieves the simulation of the relationship between CS frustration and CSRF. The following four algorithms perform the simulation process to obtain the most likely posterior probability of the CS outcome.

Algorithm 1 Calculate Posterior of CS With Respect to Frustration

Input: set cs_p, set F_{eff}
Output: cs_p
Initialization :
for each i in F_{eff} **do**
  for each j in cs **do**
    calculate posterior and replace j by posterior
  end for
end for
return return updated cs_p

Algorithm (1) evaluates the posterior probabilities of CS under the umbrella of frustration effects. It takes two sets cs_p and F_{eff} as an input while returns an updated probabilities set of CS outcome (cs_p) as an output. In addition, cs_p consists of CS outcome while F_{eff} holds the four effects of frustration. The proposed algorithm calculates the posterior of CS outcome under the influence of different frustration effects. In each iteration (under F_{eff}), the existing posterior probability is replaced by a new posterior probability. Due to this particular innovation, the proposed technique is referred to as a novel featured CS measurement algorithm. Finally, the algorithm returns an updated set of posterior probabilities with respect to all effects of frustration. These particular posterior probabilities of CS outcomes are achieved by Eq.(4) to Eq.(8).

Algorithm (2) estimates the posterior of CS outcome under the umbrella of factor mother job (CSRF). This algorithm is divided into four main iterations. In each iteration, we have 21 sub-iterations (CS outcomes). As an input, it takes cs_p (obtained from the algorithm (1)) as well as $CSRF_m$ (set of four types of mother job). It generates four sets of output as

Algorithm 2 Refining Posterior of CS With Respect to Mother Job

Input: set cs_p, set $CSRF_m$
Output: cs_s, cs_t, cs_h, cs_{home}
Initialization :
Iterations under service:
for each i in cs_p **do**
  calculate posterior and add to cs_s
end for
Iterations under teacher:
for each j in cs_p **do**
  calculate posterior and add to cs_t
end for
Iterations under health:
for each k in cs_p **do**
  calculate posterior and add to cs_h
end for
Iterations under at home:
for each m in cs_p **do**
  calculate posterior and add to cs_{home}
end for
return return cs_s, cs_t, cs_h, cs_{home},

cs_s (CS outcome with respect to mother job as a services), cs_t (CS outcome with respect to mother job as a teacher), cs_h (CS outcome with respect to mother job related to health care), cs_{home}(CS outcome with respect to mother job = at home). Firstly, this algorithm calculates posterior probabilities of CS with respect to mother job. Secondly, it divides it into four sets. The algorithm uses Eq. (9) to Eq.(13) to perform these calculations.

Algorithm 3 Refining Posterior of CS With Respect to Urban

Input: sets cs_s, cs_t, cs_h, cs_{home}
Output: sets cs_{sU}, cs_{tU}, cs_{hU}, cs_{homeU}
Initialization :
Iterations under urban:
for each i in cs_s **do**
  calculate posterior and add to cs_{sU}
end for
for each j in cs_t **do**
  calculate posterior and add to cs_{tU}
end for
for each k in cs_h **do**
  calculate posterior and add to cs_{hU}
end for
for each m in cs_{home} **do**
  calculate posterior and add to cs_{homeU}
end for
return return cs_{sU}, cs_{tU}, cs_{hU}, cs_{homeU}

A student can keep one or more exposure values. Therefore, the factor exposure is divided into two algorithms; the first algorithm evaluates CS outcome with respect to frustration and Mother job, and the second algorithm achieves posterior probabilities of CS outcomes under the influence of urban and rural layers of the CSRF. Thus, the second algorithm estimates the posterior of CS outcome (obtained from the algorithm (2)) under the urban influence. The algorithm (3) has four input sets as cs_s, cs_t, cs_h and cs_{home} while four new sets as cs_{sU}, cs_{tU}, cs_{hU} and cs_{homeU} (CS outcome with respect to frustration effects, mother jobs and urban). Each set of input is investigated under the urban impact. Moreover, it has four main iterations (as an input sets), and in each iteration, the algorithm has 21 sub-iterations. Therefore, the proposed technique achieves four updated sets of CS outcomes with respect to urban factor. The output of this particular algorithm shows a portion of our goal. It also manifests four types of posterior probabilities for four distinct individuals having a different kind of mother job.

Algorithm 4 Refining Posterior of CS With Respect to Rural

Input: sets cs_s, cs_t, cs_h, cs_{home}
Output: sets cs_{sR}, cs_{tR}, cs_{hR}, cs_{homeR}
Initialization :
Iterations under rural:
for each i in cs_s **do**
  calculate posterior and add to cs_{hR}
end for
for each j in cs_t **do**
  calculate posterior and add to cs_{hR}
end for
for each k in cs_h **do**
  calculate posterior and add to cs_{hR}
end for
for each m in cs_{home} **do**
  calculate posterior and add to cs_{homeR}
end for
return return cs_{sR}, cs_{tR}, cs_{hR}, cs_{homeR}

The last algorithm (4) uses the updated sets of the student's CS outcomes which are obtained from the algorithm (2). Algorithm (4) is not linked with the algorithm (3) but the same types of iterations performed under the umbrella of exposure (while exposure = rural). Therefore, we obtain four separate sets of posterior probabilities of CS outcomes with respect to frustration, mother jobs, and rural. Moreover, Fig. (1) represents the framework of the proposed algorithms. This particular figure shows four main modules of the current approach. The first module represents prior probabilities of CS outcomes. The selection of these probabilities is based on literature studies of psychology and neurosciences. The next module represents frustration which is divided into four layers (F1 to F4). The posterior probabilities obtained with respect to the first effect are then used as a set of prior probabilities in the next iteration (and so on). Ultimately, the final posterior probabilities of CS outcomes are used as a set of prior probabilities in the third module. This particular module is referred to as CSRF (Mother job). This module provides four sets of posterior probabilities with respect to mother jobs as services, teacher, health and at home. Furthermore, each set of posterior probability is assigned to the fourth module that generates eight sets of posterior probabilities. Therefore, the technique finalizes the simulation of each component of the CS outcome with respect to every effect of frustration and CSRF.

IV. EXPERIMENT

The current work has divided results of the experiment into different sections. Firstly, the simulations of the relationships between CS and frustration are tested. Secondly, we have tested the simulation of the statistical association between CS and CSRF. To present the results of the experiments, we have divided CS outcomes (0 to 20) into five partitions while each partition has four CS outcomes. To test the posterior probabilities of CS outcomes, the partitions are named as very low (partition (1)), low (partition (2)), average (partition (3)), good (partition (4)) and excellent CS outcome (partition (5)). During the experiment, each partition of CS is examined separately. The posterior probabilities of these partitions are evaluated under the influence of each effect of frustration as well as CSRF.

A. DATASET

We have examined the performance of the proposed technique using two types of datasets. One is public dataset [42] and the other dataset was created during psychological experiments to simulate the relationship between CS and Basic Human Factors [4]. The public dataset has contributed 70% (donated on $2014-11-27$) to the dataset section of the current attempt. On the other hand, the psychological experiment based dataset has provided 30% students' performance data for the proposed technique validation process. The public dataset contains information about CSRF and student's grades (G1, G2, and G3). We have selected G3 as the student's CS outcome which has a specific range (0 to 20). The proposed technique simulates the correlation between CSRF and G3 while ignoring the other attributes of the dataset.

Besides, the psychological experiment based dataset is used to test the posterior probabilities of CS under the umbrella of frustration. In this dataset, the target attribute G3 (CS) has a strong correlation with other student's characteristics and family-related attributes. The psychologist used specific techniques (told the student that you have failed or you can fail) to increase the frustration of a student. During frustration, different mathematical puzzles were used to check the influence of the effects of frustration. After the solution of the puzzles the student was informed about the nature of the experiment (to decrease his/her frustration level), and then a new set of puzzles were used to check his/her skills without frustration. Finally, the psychological experiment based dataset was added to the public dataset. This particular dataset has covered the deficiency of the missing factors as frustration effects. Thus, the second dataset is used for the simulation of the relationship between frustration effects and CS. Appendix shows a sample of the

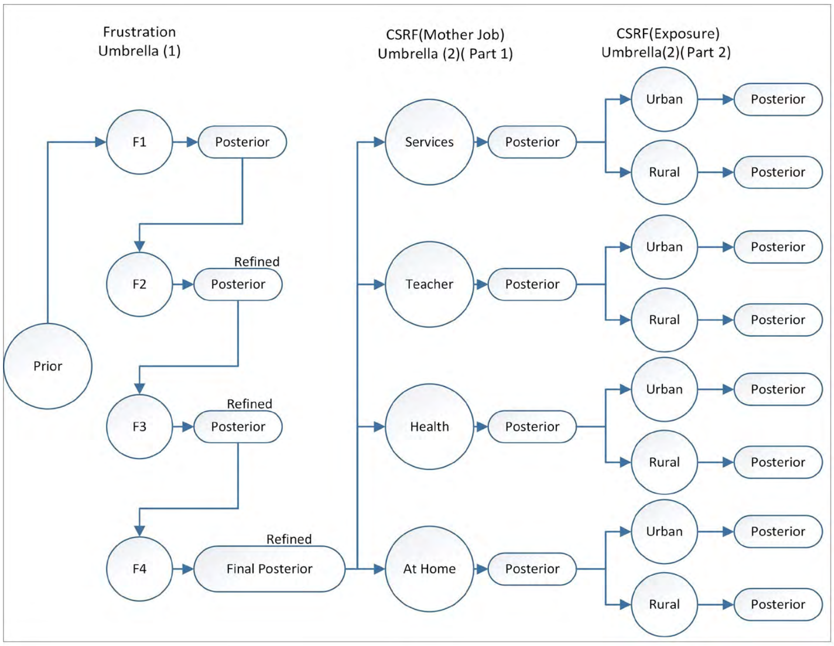

FIGURE 1. Represents two umbrellas (frustration and CSRF) with multiple layers. The posterior of CS outcome are refined in each layer of the proposed technique.

psychological experiment based dataset. The first column of the table illustrates the related attributes of the dataset while the rest of the columns manifest the values of these particular attributes.

B. SIMULATION RESULTS

The data analysis has illustrated that the posterior probabilities of very low CS outcomes, low, as well as average partition CS outcomes, are increasing with frustration effects. The posterior probabilities of the last two partitions (good and excellent CS outcomes) are decreasing due to the negative relationship between frustration and CS. Thus, in the following sections, we present that the proposed technique simulated the negative correlation between CS and frustration effects.

1) SIMULATION RESULTS OF FRUSTRATION

The posterior probabilities of very low, low and average partition of CS outcomes have a positive relationship with frustration effects. Evaluating the posterior probabilities of the first three partitions are increasing under the umbrella of frustration (effect (1) to effect (4)). It shows that the frustration effects have decreased the cognitive outcome (CS) of an individual by increasing the posterior probabilities of very low, low and average partition of the CS outcomes. If the posterior probabilities of these partitions have increased, then it shows that the expected values of lower CS outcomes have also increased. Also, the rate of increase in the posterior probabilities of partition (1) is higher than the partition (2) and (3) while the rate of increase (in the posterior probabilities) of the partition (2) is comparatively higher than partition (3). Fig (2) shows the simulation of the relationship between partition (1) and frustration effects. The different line graphs illustrate multiple CS outcomes (Mention the line number). On the hand, the x-axis shows frustration effects (F1 to F4) while the y-axis illustrates the posterior probabilities of partition (1). The rate of increase in posterior probabilities is considerably high.

Moreover, the obtained posterior probabilities (under the influence of F1) of CS outcomes are used as a set of prior probabilities in the next iteration of frustration effects (F2). The immense increase under the influence of layer F4 shows a combination of the adverse effects of F1 to F4. Figure (3) and (4) illustrated the case studies of partition (2) and partition (3) respectively. The particular rate of increase in the posterior probabilities of partition (2) is slower as compared to the partition (1). In each iteration, the negative action of frustration is increasing. These results show that the adverse effect of the last iteration is more severe as compared to the other three effects of frustrations (F1 to F3). Moreover, for every component of CS (0 to 20), the proposed method produced posterior probabilities. Thus, the frustration effects have increased the probabilities of those CS outcomes which is comparatively low. e.g., increasing the probabilities of 4 (CS outcome) as compared to greater than 4.

On the other hand, Fig. (5) and (6) shows the rate of change in the posterior probabilities of partition (4) (good CS outcomes) and partition (5) (excellent CS outcomes). Here the posterior probabilities of CS are decreased because the frustration has a negative relationship with good and excellent CS outcome. This negative relationship is shown in the form of posterior probabilities of CS outcomes. Moreover, due to the adverse statistical association of frustration and CS, the posterior probabilities of partition (4) and (5) are decreasing. The frustration effects have decreased the probabilities of partition (4) and partition (5). Therefore, Fig. (5) and (6) showed that the probabilities of CS outcome are decreasing.

2) SIMULATION RESULTS OF CSRF

Furthermore, the next umbrella is CSRF that contains two CSRF's factors, i.e., mother job, and exposure. As we discussed earlier that we divided mother job into four outcome variables (1: healthcare related, 2: teacher, 3: services, e.g., administrative or police, 4: at_home). Moreover, the variable exposure is divided into two outcomes (urban and rural). The technique evaluated the posterior probabilities under the impact of the mother job. The positive and negative relationships (between CS and mother jobs) are simulated that are shown in Fig. (7), (8), (9) and (10). According to the simulation results, the first dominant mother job is services. In addition, the mother job as a teacher and health-related worker play an average role during cognitive tasks (job interviews or written examinations). The results show that the mother job as the at_home is the least effective job during cognitive tasks.

These mother jobs are referred to as layers of the CSRF umbrella. We have studied 4 cases under each mother job. The posterior probabilities under the influence of services are shown in Fig. (7). The most probable outcome of CS is between 18 and 20 (the most probable outcome of CS under services) which is the highest outcome as compared to the rest of the three cases. On the other hand, the most probable outcome of CS under the layer of the teacher is 15. These results are shown in Fig. (8). Thus, the most likely outcome for a student is 15 under the influence of a teacher. Furthermore, in Fig. (9), the third case study illustrated the outcome of CS under the influence of health. The most likely posterior probability of CS is between 12 and 14 (CS outcomes) which is comparatively lower than the teacher and services. The last case study of a mother job is at_home that is shown in Fig. (10). It also shows that the probable outcome of CS under the at_home layer is from 6 to 8 which is the lowest outcome for a student as compared to the rest of three mother jobs.

Moreover, we have evaluated the posterior probabilities of CS outcome under the influence of urban and rural exposure. The hypothesis revealed that the CS of an individual belong to urban, is comparatively higher than the individual belongs to rural areas. The results of this case study are shown in Fig. (11). The red line shows the CS outcome concerning rural while the blue line graph illustrates the result with respect to urban factor. There are two most likely posterior probabilities: one is between 15 and 20 while the other most probable CS outcome is nearly 10. Under the influence of the urban factor, the most probable outcome of CS lies between 15 to 20. This particular urban factor has higher posterior probability (for higher CS outcome) as compared to rural factor. So, the proposed technique obtained the primary objective by simulating the correlation between CS, frustration effects and CSRF.

In the experiment section, we also have presented the comparison between the improved posterior probabilities and the initial prior probabilities. Fig. (12) shows the particular difference between the prior and posterior probabilities. The blue line graph illustrates prior while the red line graph shows the obtained posterior probabilities. The x-axis manifests the number of CS outcomes while the y-axis shows the probabilities level. This particular figure presents the significant difference in the improved posterior and prior probabilities. Fig. (13), presents the prior and posterior probabilities of CS concerning the variable exposure. Besides, Fig. (12) and (13) show the probabilities of different outcomes, but we need to demonstrate the difference between prior and improved posterior probabilities. Fig. (14) and (15) illustrate two case studies of the partition (1) and partition (5). The blue line graphs show prior probabilities while the red dotted line graphs manifest posterior probabilities of CS outcomes. These particular figures have illustrated a significant improvement in the posterior probabilities of the student's CS. The growth in the posterior probabilities is dependent on the conditional, joint and marginal probabilities. By obtaining such results, the measurements of precision and recall are not compulsory because the achieved results show a significant difference in the posterior and prior probabilities.

C. ACCURACY ANALYSIS

The proposed technique has evaluated the posterior probabilities of the CS outcomes according to the two umbrellas (frustration effects and CSRF). In each iteration, our method produced a set of 21 posterior probabilities. The technique selected outcomes with the highest posterior probability (in rest of the 20 posterior probabilities). Now, to find the accuracy of the proposed technique, we have measured precision, recall and an F1 score of the prediction results. During this process, the partitions of the CS outcomes are represented by VLP (very low CS partition), LP (low CS partition), AVP (average CS partition), GP (good CS partition), and EP (excellent CS partition) respectively. The results of accuracy (preciseness) are given in Table (1). The measures have shown that the proposed technique outperformed the existing technique because it has achieved a significant precision, recall, and F1 score values. Thus, firstly, we have measured the accuracy of the VLP which shows that precision is 0.959, recall is 0.892, an F1 score is 0.9242 while the accuracy is 0.949. Thus, we have achieved significant accuracies on these particular measures. Secondly, the technique has obtained excellent accuracy on the LP partitions (i.e., 0.973 precision, 0.917 recall, 0.9441 F1 scores, and 0.967 accuracy measure). These results show that the proposed technique outperformed the prior approaches of CS measurement. Thirdly, we have analyzed the results of AVP partition of the model. During this process, the proposed technique has achieved excellent accuracy values for AVP. The particular precision, recall, F1 score, and accuracy have scored 0.948, 0.909, 0.928 and

TABLE 1. Qualitative measures.

CS Partitions	Precision	Recall	F1 Score	Accuracy
VLP	0.959	0.892	0.9242	0.949
LP	0.973	0.917	0.9441	0.967
AVP	0.948	0.909	0.928	0.943
GP	0.958	0.933	0.9453	0.941
EP	0.987	0.898	0.9403	0.957

0.943 respectively. The results are entirely satisfactory for AVP partition of the CS outcomes. Furthermore, we have investigated the accuracy of the fourth partition of CS outcomes which is represented by GP. The accuracy results of GP are significant because it achieved a good precision, recall, F1 score and accuracy measures values. More particularly, the current approach obtained 0.958 precision, 0.933 recall, 0.9453 F1 scores as well as 0.941 accuracy value. Finally, the technique was examined for the accuracy of the last partition of CS (which is referred to as EP). The performance measures show that the proposed technique has accurately predicted CS in EP partition of the model. Thus, it has achieved significant results for precision (0.987), recall (0.898), F1 score (0.9403), and accuracy measure (0.957).

D. COMPARATIVE ANALYSIS

We conducted the comparison of the proposed technique with three competitive methods as Automatic Discovery of Cognitive Skills (ADCS) [1], Biologically Inspired Cognitive Skills Measurement Approach (BICSM approach) [4] and Gauss-Newton Algorithm (GNA). The ADCS has proposed a technique that discovered the CS of a student by using students' performance data (solved exercises). In this technique, latent skills have attached to each exercise. These skills are increased by correctly solving the exercise. Moreover, prior probabilities of students have calculated by the Weighted Chinese Restaurant Process. Comparatively, our technique is novel and differ from ADCS. The proposed method extended the motivation of ADCS towards CS measurement during job interviews and written examinations. The ADCS is insignificant in considering those student's characteristics that can affect CS positively or negatively. Therefore, the proposed technique is novel because it simulates the nonlinear relationship between frustration effects and CSRF. Such factors can efficiently change the level of CS because these are common characteristics that play an active role in cognitive tasks as interviewing or attempting a written examination. The second innovation is to quantize CS, frustration effects and CSRF of a student. As according to ADCS, the skills of a student have no specific limits while the proposed method partitioned CS outcomes into five groups (very low CS, low CS, average CS, good CS, and excellent CS outcomes).

On the other hand, the proposed technique accurately quantized CS and further split it into 21 periodic discrete outcomes. By using this unique range, we can measure the accuracy of the proposed prediction technique. We achieved a component-wise posterior probability model by splitting the range of CS into periodic values that also ensured significant prediction accuracy. The proposed technique has two umbrellas (along with the layers) as frustration and CSRF. The posterior probabilities of CS outcomes are refined under these umbrellas by our method. To improve the accuracy of the proposed technique, we evaluated the posterior probabilities of each outcome of CS with respect to frustration effects and CSRF.

Furthermore, we compared the proposed CS measurement technique with another competitive method known as BICSM approach. This competitive method proposed CS measurement algorithm that has simulated the nonlinear relationship between CS and Basic Human Factors (BHF) (aging, infection, emotions, awareness, personality, education, and experience). This algorithm was split into three sub-algorithms, (1) estimate BHF factor values, (2) validate the estimated values of BHF, and (3) measure CS by using the estimated values of BHF. This method has proposed a unique range of CS (0 to 10 on the Likert scale). In addition, it has not quantized (explicitly) the particular range of CS. It has also proposed specific domains and ranges for BHF. Therefore, the nonlinear least square method was used to estimate the unknown parameters. Estimations of parameters have modulated the relationship between CS and BHF. The proposed technique is novel, efficient and more detailed as compared to the BICM approach. It can measure CS during interviews and written examinations by targeting specific student's characteristics that have a close relationship with CS. Additionally, BICSM approach seemed to be ambiguous because it has many BHF. Factor-wise quantization and simulation of the relationship between CS and the large set of BHF can compromise the accuracy of the prediction method. Besides, our technique is novel and accurate due to the following features.

- The proposed technique refined the posterior probabilities of CS (21 outcomes) with respect to frustration effects.
- It received the refined set of posterior probabilities from frustration umbrella and then re-refined it with respect to each layer of CSRF umbrella.

Therefore, the proposed technique evaluated the posterior probabilities of CS with respect to each layer of the two umbrellas. Furthermore, the proposed method is tested on the dataset which was used to validate the accuracy of BICSM approach. The results show that the proposed method achieved a significant accuracy as compared to BICSM approach. Table (2) represents these particular accuracy results which manifest that the proposed technique achieved

TABLE 2. Accuracy comparison.

	MFE	MAD	TS
BICSM Approach	0.091(91%)	1.731	1.0
Our technique	0.097(97%)	1.621	-1.3

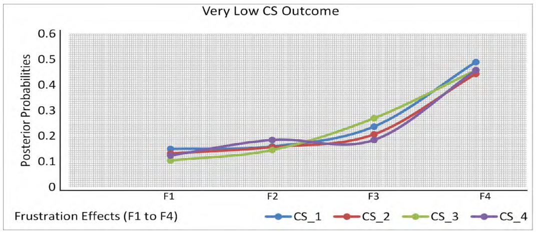

FIGURE 2. Illustrates the posterior of partition (1). It shows that the posterior probabilities is increasing with with frustration effects (F1 to F4).

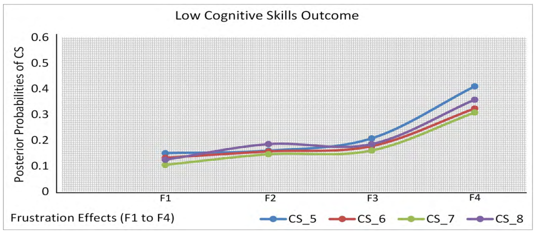

FIGURE 3. Shows the results the results of partition (2). Its also illustrates that the posterior probabilities of CS outcomes increasing with frustration effects but the rate of increase is lower than the partition (1).

a significant accuracy on Mean Forecast Error (MFE), Mean Absolute Deviation (MAD) and Tracking Signal (TS) measures. So, the proposed method obtained MFE = 0.097 which represents that the technique accuracy is 97% because MFE error is 3%. Finally, we obtained MAD = 1.621 and TS=−1.3 which show that our technique is comparatively significant.

Furthermore, the proposed technique is compared with the Gauss-Newton Method (GNM). The GNM method is designed to fit a novel mathematical model to the data. The basic function of GNM is to simulate the nonlinear correlation between dependent (CS outcomes) and independent (frustration effects and CSRF) variables ($y = f(x, \theta) + \epsilon$) using the nonlinear least square technique. The mathematical model for the simulation of the correlation between CS, CSRF and frustration effects are nonlinear in parameters. Thus, here we briefly explore the GNM technique for further error analyses. In GNM, we transform the nonlinear model into the locally linear model. Therefore, we have the initial model as follow.

$$cs_i = f(F^{(i)}_{csrf}, \theta) + \epsilon_i \tag{14}$$

In Eq.(14), the error $\epsilon_i = cs_i - (F^{(i)}_{csrf}, \theta)$ while cs_i represents CS outcome, F_{csrf} shows frustration effects and CSRF. Furthermore, we minimized the error with respect to θ ($\theta^{Min}\{\sum_{i=1}^{N} \epsilon_i\}$). To initiate the iteration of GNM, we need the

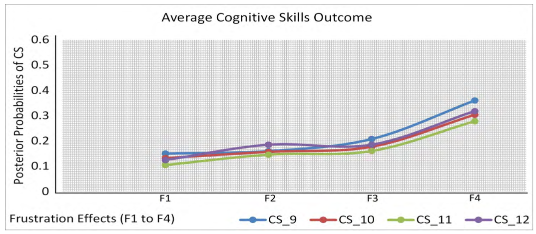

FIGURE 4. Represents the posterior probabilities of partition (3). Posterior of CS outcomes increasing but the rate increase is lower as compare to partition (2).

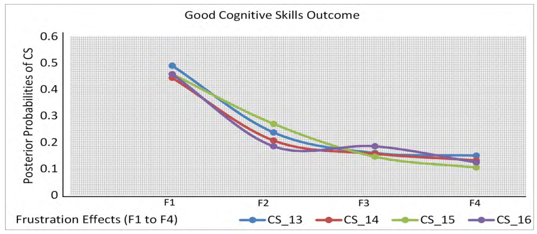

FIGURE 5. Illustrates the results of partition (4). The posterior probabilities is decreasing due to negative effects of frustration effects.

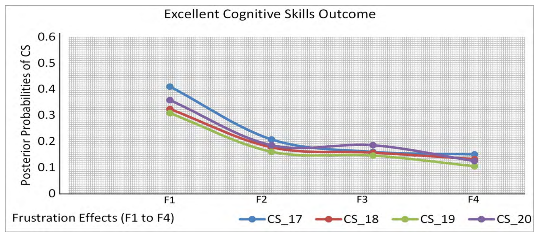

FIGURE 6. Shows that the posterior probabilities of CS outcome is decreasing with frustration effects but the rate of decrease is higher as compare to partition (4).

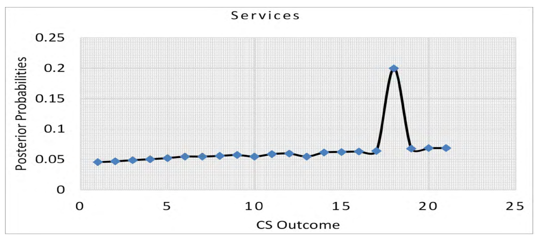

FIGURE 7. Illustrates that under the influence of mother job as services, the propose method achieved higher posterior probabilities for [16 to 18].

initial guess for the parameters θ. Thus, lets say $\bar{\theta}$ = initial guess solution. Now, to transform the nonlinear parameter model to linear model, we use Taylor series. So, we obtain the following model.

$$cs = f(F_{csrf}, \bar{\theta}) + (\frac{\partial f}{\partial \theta_1})_{\theta=\bar{\theta}} \Delta\theta_1 + \ldots + (\frac{\partial f}{\partial \theta_m})_{\theta=\bar{\theta}} \Delta\theta_m + \epsilon \tag{15}$$

Eq. (15) shows that Taylor series has two parameters (θ and $\bar{\theta}$). The original function (Eq. 14) was nonlinear while we have linearized it by transforming to Taylor series.

TABLE 3. Psychological experiment based dataset sample.

sex	F	M	F	F	M	M	M	F	M	F	F
age	20	21	21	18	16	17	19	19	15	20	17
address	R	R	R	R	R	R	R	R	R	R	R
famsize	LE3	GT3	LE3	GT3	GT3	LE3	LE3	GT3	LE3	LE3	GT3
Medu	1	2	1	4	3	4	2	4	3	3	4
Fedu	3	3	1	2	2	3	2	3	2	3	3
Mjob	at_{home}	at_{home}	at_{home}	health	other	services	other	other	services	other	teacher
Fjob	services	services	services	services	services	services	teacher	teacher	teacher	teacher	teacher
reason	other	other	other	home	home	repu	home	home	home	home	repu
guardian	father	father	mother	father	father	uncle	mother	mother	brother	mother	mother
traveltime	2	2	1	0	1	1	1	1	1	1	1
studytime	3	4	2	4	2	2	3	2	2	2	2
famsup	yes	yes	no	yes	no	yes	no	yes	yes	yes	yes
paid	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
activities	yes	no	no	yes	no	yes	no	yes	no	yes	no
nursery	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes
higher	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
internet	no	yes	yes	yes	no	yes	yes	no	yes	yes	yes
freetime	3	3	3	2	3	4	4	1	2	5	3
health	3	3	3	5	5	5	3	1	1	5	2
absences	6	4	10	2	4	10	0	6	0	0	0
Pstatus	T	T	T	T	T	T	T	T	T	T	T
famsup	no	no	no	yes	yes	yes	no	yes	yes	yes	yes
famrel	4	4	4	3	4	4	4	4	4	5	3
health	4	3	3	4	5	4	3	1	1	5	2
absences	6	4	10	2	4	10	0	6	0	0	0
aggression	yes	yes	no	no	yes	yes	yes	yes	no	no	yes
giving up	yes	yes	yes	no	no	yes	yes	yes	yes	no	no
LSC	yes	no	no	no	yes	yes	yes	no	no	no	yes
stress	no	yes	no	yes	yes	yes	no	yes	no	yes	yes
G3 (CS)	8	7	12	15	13	6	7	6	10	13	5

Appendix A shows a sample from the psychological experiment based dataset. The first column of the table illustrates all the attributes of the dataset while the rest of the columns manifest the values of these particular attributes

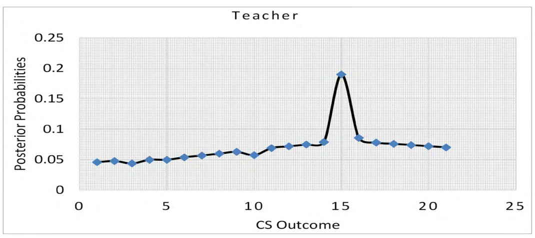

FIGURE 8. Manifests that most probable outcome of CS (under the influence of layer teacher) is 15.

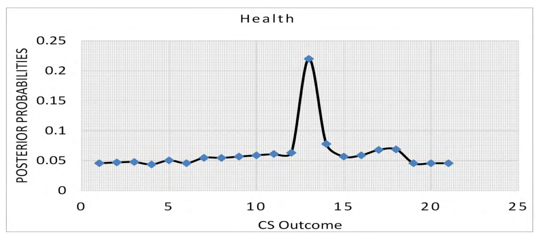

FIGURE 9. Presents that the most probable outcome of CS under the influence of health (mother job) is nearly 14 (between 12 and 14).

Now, we can technically compare the proposed method with GNM. Comparatively, choosing GNM model does not guarantee an error-free solution for CS measurement. Therefore, the accuracy and flexibility of the mathematical model can be compromised due to the following errors.

- Deviation in the values of independent variables.
- Error in the measurement of the dependent variable.
- Accuracy issues in the proposed novel equation.

The three types of errors are common during GNM model process; therefore, these issues motivated us to develop a novel technique for CS measurement. On the other hand, we achieved a precise probabilistic model which iteratively refined (calculated) the posterior probabilities of

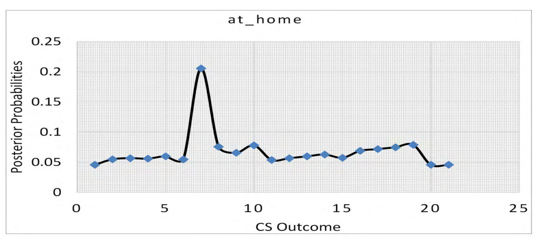

FIGURE 10. This figure manifests that the most probable outcome of CS is between 6 to 8.

the CS outcomes. It also ensured prediction accuracy by dividing CS into periodic discrete outcomes (0 to 20) with a period of 1. Besides, the proposed technique focused on

IEEE Access

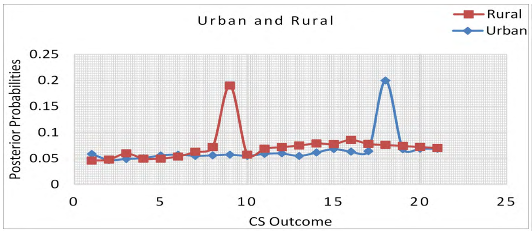

FIGURE 11. Shows that under the rural layer, we achieved higher posterior for lower CS outcomes.

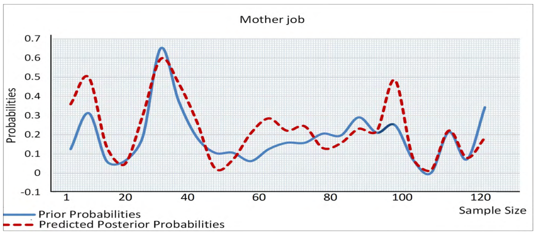

FIGURE 12. Illustrates the enhance probabilities of CS outcome under the layer of mother job. This shows the difference between prior and posterior probabilities obtained during experiment.

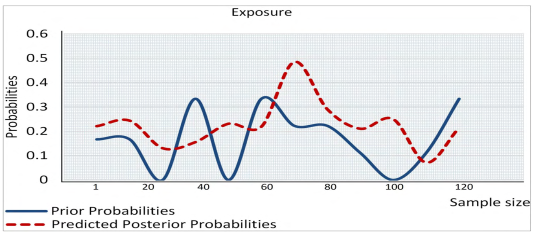

FIGURE 13. It also manifests the particular enhancement in prior probabilities under the layer of exposure. The red dotted line graph show enhanced posterior probabilities of CS outcome.

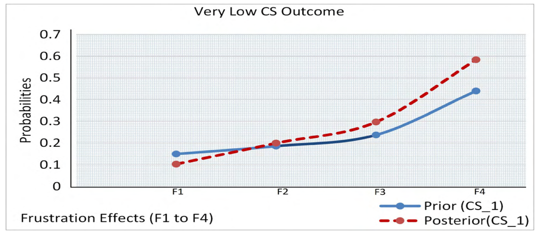

FIGURE 14. Shows the probabilities of CS outcome belong to the partition (1). The enhanced probabilities is shows in dotted line graph.

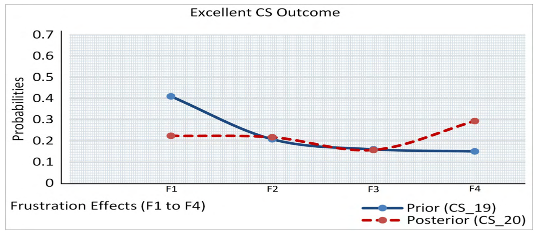

FIGURE 15. It manifest posterior of CS outcome from the partition (5). It shows enhancement in prior. The dotted line graph is posterior probabilities which is also referred as enhanced probabilities of CS outcome.

accuracy by dividing frustration umbrella into four layers while CSRF umbrella into six layers. In each layer, the technique evaluated the posterior probability of every CS outcome. Moreover, the proposed technique can operate on large and small datasets which provided a dynamic environment to overcome the prediction errors.

V. DISCUSSION

In the current attempt, we presented a novel technique to predict a specific level of CS by simulating the nonlinear relationship between CSRF, frustration, and CS. The first contribution of the proposed technique is to select a particular range for CS (0 to 20) because it ensured accurate quantization. The chosen range made it easy for CS to be broken into further components. Thus, during quantization, CS range is divided into 21 periodic discrete outcomes with a period of 1. The second contribution is to simulate the relationship between CS and frustration. For the unbiased estimation of each CS outcome, the proposed technique divided frustration into four effects (aggression, giving up, loss of self-confidence and stress). The frustration is referred to as an umbrella while the effects of frustration are referred to as layers of the umbrella. The technique calculated the posterior probabilities of CS outcomes under the layers of frustration umbrella. As a third contribution, the method simulated the nonlinear relationship between CS and CSRF. The CSRF is divided into two variables; (1) mother jobs and (2) exposure. Furthermore, mother jobs are split into four outcomes (services, teacher, health, and at home) while exposure has two outcomes (urban and rural). The CSRF referred to as the second umbrella while the factors of CSRF are called as layers of the umbrella. To automate the measurement of CS, we re-estimated the posterior probabilities of CS outcomes under the umbrellas of CSRF. During the extensive experiments, we tested the proposed method on two datasets. The results of the experiment showed that the proposed technique efficiently simulated the relationship between CS, CSRF, and frustration. This indicates that the proposed technique accurately evaluated the posterior probabilities of CS outcomes (see Fig. 2 to Fig. 15). Finally, we concluded our work by presenting the obtained prediction accuracy (see Table. (1)) and comparison with competitive methods.

VI. CONCLUSION

The current attempt presented a novel prediction technique that used the nonlinear correlation between CS, frustration, and CSRF for the estimation of student's performance. Firstly, the technique proposed a unique range of CS (0 to 20) for the accurate quantization of the student's skills. This range is further split into 21 periodic discrete outcomes. Secondly, frustration is divided into four effects that are referred to as layers of the umbrella. Thirdly, it divided CSRF into two variables; (1) mother jobs (services, teacher, health and at home) and (2) exposure (urban and rural).

CSRF is also called as a six-layered umbrella. To accurately estimate the CS of a student, the technique iteratively calculated the posterior probability of each outcome of CS under the frustration umbrella. Moreover, the proposed technique re-estimated the achieved posterior probabilities under the influence of the CSRF umbrella. Finally, the current method is tested on two datasets that showed surprising accuracy results.

While working on the CS measurement technique, we have reported some limitations due to the measurement of frustration effects. We need a novel technique to simulate the relationship between CS and frustration using Electroencephalography (EEG) signals [43], [44]. Such factors need to be quantized and supported by EEG signal patterns because tracking the emotional status of an individual is essential for CS measurement methods. Furthermore, the proposed technique may produce a different result by replacing the Bayesian inference method with other methods (e.g., Naive Bayes classifier). (We have performed a series of experiments and then evaluated the technique on average performance in term of state-of-the-art measures.)

APPENDIX
SAMPLE OF DATA

See Table 3.

REFERENCES

[1] R. V. Lindsey, M. Khajah, and M. C. Mozer, "Automatic discovery of cognitive skills to improve the prediction of student learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1386–1394.

[2] M. Bierzynska *et al.*, "Effect of frustration on brain activation pattern in subjects with different temperament," *Frontiers Psychol.*, vol. 6, p. 1989, Jan. 2016.

[3] J. Cai and A. Stoyanov, "Population aging and comparative advantage," *J. Int. Econ.*, vol. 102, pp. 1–21, Sep. 2016.

[4] S. Ahmad, K. Li, H. A. I. Eddine, and M. I. Khan, "A biologically inspired cognitive skills measurement approach," *Biologically Inspired Cogn. Archit.*, vol. 24, pp. 35–46, Apr. 2018.

[5] A. J. Egalite, "How family background influences student achievement," *Educ. Next*, vol. 16, no. 2, pp. 70–78, 2016.

[6] R. V. Lindsey, J. D. Shroyer, H. Pashler, and M. C. Mozer, "Improving students' long-term knowledge retention through personalized review," *Psychol. Sci.*, vol. 25, no. 3, pp. 639–647, 2014.

[7] K. H. Wilson *et al.*, "Estimating student proficiency: Deep learning is not the panacea," in *Proc. Neural Inf. Process. Syst., Workshop Mach. Learn. Educ.*, 2016, p. 3.

[8] K. R. Koedinger, E. A. McLaughlin, and J. C. Stamper, "Automated student model improvement," in *Proc. Int. Educ. Data Mining Soc.*, 2012, pp. 1–8.

[9] B. Qiao, S. Liu, H. Zeng, X. Li, and B. Dai. (2015). "Limitation of the least square method in the evaluation of dimension of fractal Brownian motions." [Online]. Available: https://arxiv.org/abs/1507.03250

[10] C. Ugrinowitsch, G. W. Fellingham, and M. D. Ricard, "Limitations of ordinary least squares models in analyzing repeated measures data," *Med. Sci. Sports Exerc.*, vol. 36, no. 12, pp. 2144–2148, 2004.

[11] J. Chen, K. Li, Z. Tang, K. Bilal, and K. Li, "A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment," *IEEE Access*, vol. 4, pp. 1767–1783, 2016.

[12] S. Cheng, M.-T. Martinez-Ingles, D. P. Gaillot, J.-M. Molina-Garcia-Pardo, M. Liénard, and P. Degauque, "Performance of a novel automatic identification algorithm for the clustering of radio channel parameters," *IEEE Access*, vol. 3, pp. 2252–2259, 2015.

[13] L. Zuo, L. Shu, S. Dong, C. Zhu, and T. Hara, "A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing," *IEEE Access*, vol. 3, pp. 2687–2699, 2015.

[14] Y.-H. Chou, S.-Y. Kuo, C.-Y. Chen, and H.-C. Chao, "A rule-based dynamic decision-making stock trading system based on quantum-inspired tabu search algorithm," *IEEE Access*, vol. 2, pp. 883–896, 2014.

[15] P.-H. Kuo, T.-H. S. Li, Y.-F. Ho, and C.-J. Lin, "Development of an automatic emotional music accompaniment system by fuzzy logic and adaptive partition evolutionary genetic algorithm," *IEEE Access*, vol. 3, pp. 815–824, 2015.

[16] M. Filippone and M. Girolami, "Pseudo-marginal Bayesian inference for gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2214–2226, Nov. 2014.

[17] Z. Liu, V. Pataranutaporn, J. Ocumpaugh, and R. Baker, "Sequences of frustration and confusion, and learning," in *Proc. Educ. Data Mining*, 2013, pp. 1–7.

[18] N. Hara, "Student distress in a Web-based distance education course," *Inf., Commun. Soc.*, vol. 3, no. 4, pp. 557–579, 2000.

[19] J. Wilde, "The relationship between frustration intolerance and academic achievement in college," *Int. J. Higher Educ.*, vol. 1, no. 2, p. 1, 2012.

[20] M. S. Khan, G. Mujtaba, M. A. Al-garadi, N. H. Friday, A. Waqas, and F. R. Qasmi, "Multi-soft sets-based decision making using rank and fix valued attributes," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, 2018, pp. 1–11.

[21] M. S. Khan, M. A. Al-Garadi, A. W. A. Wahab, and T. Herawan, "An alternative data filling approach for prediction of missing data in soft sets (adfis)," *SpringerPlus*, vol. 5, no. 1, p. 1348, 2016.

[22] M. Khajah, R. V. Lindsey, and M. C. Mozer. (2016). "How deep is knowledge tracing?" [Online]. Available: https://arxiv.org/abs/1604.02416

[23] M. S. Khan, A. W. A. Wahab, T. Herawan, G. M. Shaikh, S. Danjuma, and M. A. Al-Garadi, "Virtual community detection through the association between prime nodes in online social networks and its application to ranking algorithms," *IEEE Access*, vol. 4, pp. 9614–9624, 2016.

[24] J. Hughes and O.-M. Kwok, "Influence of student–teacher and parent–teacher relationships on lower achieving readers' engagement and achievement in the primary grades," *J. Educ. Psychol.*, vol. 99, no. 1, pp. 39–51, 2007.

[25] M. Talanov *et al.*, "Emotional simulations and depression diagnostics," *Biologically Inspired Cogn. Archit.*, vol. 18, pp. 41–50, Oct. 2016.

[26] A. V. Samsonovich, "Emotional biologically inspired cognitive architecture," *Biologically Inspired Cogn. Archit.*, vol. 6, pp. 109–125, Oct. 2013.

[27] J. S. Lavelle, "Is a modular cognitive architecture compatible with the direct perception of mental states?" *Consciousness Cogn.*, vol. 36, pp. 508–518, Nov. 2015.

[28] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cogn. Syst. Res.*, vol. 10, no. 2, pp. 141–160, 2009.

[29] R. Ichise, "An analysis of the CHC model for comparing cognitive architectures," *Procedia Comput. Sci.*, vol. 88, pp. 239–244, Jan. 2016.

[30] N. M. Yusoff, R. F. Ahmad, C. Guillet, A. S. Malik, N. M. Saad, and F. Mérienne, "Selection of measurement method for detection of driver visual cognitive distraction: A review," *IEEE Access*, vol. 5, pp. 22844–22854, 2017.

[31] R. Fernandez and R. W. Picard, "Signal processing for recognition of human frustration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 6, May 1998, pp. 3773–3776.

[32] X. Guan, C.-T. Li, and Y. Guan, "Matrix factorization with rating completion: An enhanced SVD model for collaborative filtering recommender systems," *IEEE Access*, vol. 5, pp. 27668–27678, 2017.

[33] P. Zhang, X. Zhou, P. Pelliccione, and H. Leung, "RBF-MLMR: A multi-label metamorphic relation prediction approach using RBF neural network," *IEEE Access*, vol. 5, pp. 21791–21805, 2017.

[34] W. Shi *et al.*, "Temporal dynamic matrix factorization for missing data prediction in large scale coevolving time series," *IEEE Access*, vol. 4, pp. 6719–6732, 2016.

[35] W. Li, J. Wang, K. Chen, and W. Li, "An intelligent cognition method of human health states based on a variant knowledge granularity feedback mechanism," *IEEE Access*, vol. 5, pp. 19570–19580, 2017.

[36] S. Haykin, J. M. Fuster, D. Findlay, and S. Feng, "Cognitive risk control for physical systems," *IEEE Access*, vol. 5, pp. 14664–14679, 2017.

[37] C. W. Cotman, N. C. Berchtold, and L.-A. Christie, "Exercise builds brain health: Key roles of growth factor cascades and inflammation," *Trends Neurosci.*, vol. 30, no. 9, pp. 464–472, 2007.

[38] J. B. Klein, R. H. Jacobs, and M. A. Reinecke, "Cognitive-behavioral therapy for adolescent depression: A meta-analytic investigation of changes in effect-size estimates," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 46, no. 11, pp. 1403–1413, 2007.

[39] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, ''Machine learning framework for the detection of mental stress at multiple levels,'' *IEEE Access*, vol. 5, pp. 13545–13556, 2017.
[40] F. Al-Shargie, T. B. Tang, and M. Kiguchi, ''Stress assessment based on decision fusion of EEG and fNIRS signals,'' *IEEE Access*, vol. 5, pp. 19889–19896, 2017.
[41] Z. Huang, C. Wu, and Z. Wang, ''Stress direction measurement based on polarization state in optical fibers using the quaternion method,'' *IEEE Photon. J.*, vol. 9, no. 6, pp. 1–11, Dec. 2017.
[42] P. Cortez and A. M. G. Silva, ''Using data mining to predict secondary school student performance,'' Inf. Syst./Algoritmi R&D Centre, Univ. Minho, Guimarães, Portugal, Tech. Rep., 2008.
[43] J. Meng *et al.*, ''Effects of soft drinks on resting state EEG and brain–computer interface performance,'' *IEEE Access*, vol. 5, pp. 18756–18764, 2017.
[44] J. Fernández-Vargas, T. V. Tarvainen, K. Kita, and W. Yu, ''Effects of using virtual reality and virtual avatar on hand motion reconstruction accuracy and brain activity,'' *IEEE Access*, vol. 5, pp. 23736–23750, 2017.

SADIQUE AHMAD received the master's degree in computer science from IMSciences, Peshawar, Pakistan, in 2013, with a focus on cognitive skills measurement. He is currently pursuing the Ph.D. degree in computer science with the Beijing Institute of Technology. He has authored four published papers and five accepted papers in peer-reviewed journals and conferences. His research interests include machine learning and more specifically cognitive skills measurement approaches.

KAN LI is currently a Professor with the School of Computer, Beijing Institute of Technology. He has authored over 50 technical papers in peer-reviewed journals and conference proceedings. His research interests include machine learning and pattern recognition.

ADNAN AMIN received the M.S. degree (Hons.) in computer science from the Institute of Management Sciences (IMSciences), Peshawar, Pakistan, in 2015, and the M.Sc. degree in computer science from the University of Peshawar, Pakistan, in 2008. He is currently pursuing the Ph.D. degree in data mining and machine learning with the Center of Excellence in Information Technology, IMSciences. From 2009 to 2011, he was with Maxwell Stamp PLC, London, as an International Consultant for Curricular and Academic Development Activities with the School of ICT, National Institute of Management and Administration, Kabul, under the Ministry of Education and World Bank Funded Project (P102573). In 2015, he joined IMSciences as a Lecturer of CS/IT. He has conducted and participated in collaborative research projects with academia and industry. He has authored or co-authored 16+ research papers, including 7+ journal articles.

SALABAT KHAN received the master's degree in computer science from International Islamic University, Islamabad, Pakistan, in 2014, with a focus on privacy preserving secure data mining. He is currently pursuing the Ph.D. degree in computer science with the Beijing Institute of Technology.