# Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review

Emmanuel Okewu, Phillip Adewole, Sanjay Misra, Rytis Maskeliunas & Robertas Damasevicius

Published online: 09 Oct 2021.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

# Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review

Emmanuel Okewu[a], Phillip Adewole[a], Sanjay Misra[b], Rytis Maskeliunas[c], and Robertas Damasevicius[d]

[a]Centre for Information Technology and Systems, University of Lagos, Lagos, Nigeria; [b]Department of Computer Science and Communication, Ostfold University College, Halden, Norway; [c]Department of Applied Informatics, Vytautas Magnus University, Kaunas, Lithuania; [d]Faculty of Applied Mathematics, Silesian University of Technology, Gliwice, Poland

### ABSTRACT

Efforts to raise the bar of higher education so as to respond to dynamic societal/industry needs have led to a number of initiatives, including artificial neural network (ANN) based educational data mining (EDM) inclusive. With ANN-based EDM, humongous amount of student data in higher institutions could be harnessed for informed academic advisory that promotes adaptive learning for purposes of student retention, student progression, and cost saving. Mining students' data optimally requires predictive data mining tool and machine learning technique like ANN. However, despite acknowledging the capability of ANN-based EDM for efficiently classifying students' learning behavior and accurately predicting students' performance, the concept has received less than commensurate attention in the literature. This seems to suggest that there are gaps and challenges confronting ANN-based EDM in higher education. In this study, we used the systematic literature review technique to gauge the pulse of researchers from the viewpoint of modeling, learning procedure, and cost function optimization using research studies. We aim to unearth the gaps and challenges with a view to offering research direction to upcoming researchers that want to make invaluable contributions to this relatively new field. We analyzed 190 studies conducted in 2010–2018. Our findings reveal that hardware challenges, training challenges, theoretical challenges, and quality concerns are the bane of ANN-based EDM in higher education and offer windows of opportunities for further research. We are optimistic that advances in research along this direction will make ANN-based EDM in higher education more visible and relevant in the quest for higher education-driven sustainable development.

## Introduction

In a bid to improve higher education and make more responsive to the needs of industry, advances in research have led to a number of pedagogical measures such

as face-to-face learning, virtual learning, blended learning, and online learning. However, none of these have taken advantage of the huge learner-related data generated during learning to enhance decision making in the education domain. In a marked departure, educational data mining is a data-based technology-enhanced pedagogical approach that leverages on data science techniques like artificial intelligence (AI), data mining, knowledge discovery in databases (KDD), and data warehouse that harnesses learning-related data for informed decision making in the learning environment. Educational data mining (EDM) is the analysis of huge sets of learner-related (Barneveld, Arnold, and Campbell 2012; Siemens et al. 2011) with the aid of methods like KDD, business intelligence, educational data mining, social network analysis, operational research, machine learning, and information visualization with the aim of informing and shaping the learners' environments (Ali et al. 2013; Fournier, Kop, and Sitlia 2011).

One of the ways to elicit unbiased and non-prejudicial information from student data for driving smart education is the use of artificial neural networks (ANN). It is purely a data-based technique that applies function approximation (regression), pattern recognition (classification), and predictive analytics to learner-related observational to predict student performance and classify student-learning behavior with a view to improving student retention, student progression (Brocardo et al. 2017), motivation (Sak, Senior, and Beaufays 2014a) and cost saving. Hence, ANN is suitable for promoting smart education, intelligent tutoring, and offering academic advising.

This study is a systematic literature review (SLR) of the application of neural networks for educational data mining in higher education, from the perspective of modeling, learning tasks/algorithms, and cost function optimization. The ultimate goal of EDM is to use information obtained for decision making that enhances student retention, student progression, and cost saving through optimal resource utilization.

For robust pattern recognition and reliable prediction services, there is a need for a model that appropriately maps the statistical functions of student data and leverages on the function approximation to offer reliable outcomes. Hence, in this study, we explore the capability of ANN to model and learn in complex and mission-critical domains like EDM.

The motivation for this study stems from the need to foster understanding of EDM domain from the perspective of ANN modeling, learning procedures, and cost function optimization which hitherto has been under-reported, but are critical and expedient for holistic understanding of EDM in higher education with respect to research and practice. This is because a review of EDM learning tasks/algorithms will put in perspective the necessary procedures, operations, and processes of realizing ANN-based EDM; while a review of cost function optimization methods will offer an opportunity to assess EDM's utility in the higher education domain. We are of the opinion that such a tripartite approach is essential for the comprehensive understanding of the

employment of neural networks in EDM landscape with respect to research and practice.

Hence, the goals of the study include i) to offer reliable intellectual guide for future researchers in neural networks and EDM to enable them identify trend of EDM research in the context of modeling, learning procedures, and cost function optimization; and ii) to develop a portrait of current knowledge in respect of existing patterns of research commitments in these aspects, and existing relevant gaps in the areas of EDM modeling, learning procedures, and cost function optimization.

In their view, Macfadyen and Dawson (Macfadyen and Dawson 2012) and few others posit that the unresolved issues and open problems of ANN-driven EDM in higher education can impede the delivery of benefits such as student retention, student progression and cost saving. For instance, significant processing and storage resources are needed to implement complex and effective software neural networks (Edwards 2015). Though many SLRs have been reported in the emerging field of EDM, none had approached it from the perspective of ANN modeling, learning tasks/algorithms, and cost function optimization as attempted in this study. We hinge our investigation on four research questions that probe the frequency of attention that each of the topics (modeling, learning procedure, and cost function optimization) has gained thus far in the literature, as well as the paradigm of research commitments in these aspects. This SLR used the model of (Kitchenham, Charters, and Kuzniarz 2015), while steps for conducting it were obtained from (Höst and Oručević-Alagić 2011).

The research questions that formed the basis of investigation in this study are as follows:

*RQ 1 – What is the trend of research interests in best-fit mapping of student data using neural network modeling, in the context of the least and most researched issues?*

*RQ 2 – What are the least, moderately and most researched learning procedures for training neural networks in pattern recognition of student data?*

*RQ3 – What are the cost function optimization methods for ascertaining stability of neural networks for classifying students learning behaviors and predicting academic performance in terms of least, moderately and most prominent?*

*RQ4 – What is the pattern of research interests in ANN-based EDM in higher education, in terms of the least and most researched open problems?*

The remainder of this article is arranged as follows. Section 2 offers a preview of background and related work, just as Section 3 elaborates on the research methodology used for the SLR. In Section 4, the limitations of the study are discussed, and report of the findings given in Section 5. In Section 6, we discuss our findings, and conclude the study in Section 7.

## Background and Related Work

This section discusses EDM, statistical learning, machine learning, and ANN. Related works on EDM in higher education are also discussed.

### *Context of Educational Data Mining*

The development of EDM is a transformation process that has faced many areas of application including educational domain modeling, learning component analysis, user profiling, user knowledge, behavior modeling, experience modeling, learning analytics, and trend analysis (Dascalu et al. 2018). As Ferguson (Ferguson 2012) opines, the use of analytics for learning is a process that evolved through some seven stages. Cooper (2012) affirmed that learning analytics draws techniques from a number of communities, including statistics, business intelligence, web analytics, operation research, artificial intelligence (AI) and data mining, information visualization, and social network analysis.

Drawing inspiration from the identified developmental stages of EDM and its community, (Chatti et al. 2014) provided a reference model. The reference model is based on four dimensions with a view to providing a systematic overview of EDM and related concepts. Also, the reference model is intended to facilitate communications among researchers in their quest to address the challenges associated with EDM evolution, against the backdrop that the technical and pedagogical issues of EDM are bound to evolve from time to time. Data Science techniques that form the basis for EDM are data mining, knowledge discovery in databases (KDD), data warehouse, and machine learning (Ali et al. 2012).

### *Machine Learning*

A framework for machine learning (ML) is statistical learning (SL) theory. SL utilizes techniques from the domains of functional analysis and statistics (Hastie, Tibshirani, and Friedman 2009; Mohri, Rostamizadeh, and Talwalkar 2012a) with a focus on the problem of establishing a predictive function based on data. Its application areas include speech recognition, computer vision, and bioinformatics (Sidhu and Caffo 2014).

Understanding and prediction are the two cardinal goals of learning and the categories of learning include supervised learning, online learning, unsupervised learning, and reinforcement learning. However, supervised learning is most prominent in SL (Poggio and Rosasco et al. 2012). Supervised learning entails learning from a set of training data. In the training process, each point is an input-output tuple, with the input mapping to an output. The aim is to infer the function that maps input to output in such a way that the learned

mapping can use future input to predict output. Supervised learning problems are categorized as either regression problems or classification problems depending on the type of output. A regression problem is one whose output involves a continuous range of values. The regression is to find the existing functional relationship between the input and output. For example, finding a functional relationship between the reading habit of a student (measured in hours) and obtained grade point average (GPA).

On the other hand, in classification problems, the output is an element from a discrete set of labels. Classification is commonly used in machine learning applications. For instance, in the field of facial recognition, the input is a picture of a person while the output label is the person's name. A large multidimensional vector is used to represent the input and the elements of the vector represent pixels in the picture. After function approximation, the learnt function is validated on a test set of data which is different from the training data.

The major problem of machine learning is overfitting. Since learning is a prediction problem, the ultimate goal is to find a function that most accurately predict output from future input even though the secondary goal is to determine a function that most closely fits the observed data. In effect, the learning resilience of a learning model is attributable to its mapping and predictive capabilities. There is therefore concern over the risk of overfitting where a function matches the data accurately but is deficient in predicting future output so that empirical risk minimization efforts would not suffer from the effects of overfitting. Overfitting indicates that a solution is unstable such that a minor change in the training data can result in high variation of the learned function. Studies have shown that once stability for the solution is guaranteed, a platform for generalization and consistency is guaranteed (Mukherjee et al. 2006; Vapnik and Chervonenkis 1971).To solve the over-fitting problem, regularization is used as it gives the problem stability (Poggio and Rosasco et al. 2012).

## Artificial Neural Network (ANN)

ANN is a data-based machine learning technique that refers to interconnected ensemble of nodes similar to the huge network of neurons in a brain. In ANN, artificial neuron is represented by a circular node while the link between the output of a neuron and the input of another is represented by an arrow. Its use implies that rather than program a system (machine) as in fuzzy logic, we train the system (machine) to learn.

ANN as a computational model finds uses in machine learning, education, computer science and other research disciplines. It is premised on a large collection of linked simple units referred to as artificial neurons, a resemblance of axons in a biological brain. Links between neurons convey activation signals

of varying strength (Schmidhuber 2015a) such that the strength of the combined incoming signals determines the level of activation of the neuron. When the neuron is sufficiently activated, the received signals are propagated to other neurons connected to it. Rather than programming such systems, they can be trained using existing examples (training or historical or observational data). They are frequent reports in the literature that trained systems typically excel in areas where the solution or feature detection is hard to express in traditional computer program. A wide range of complex tasks such as speech recognition and computer vision that are hard to solve using normal rule-based programming have been tackled using machine learning methods or intelligent technologies like neural networks, genetic algorithm, fuzzy logic, and probabilistic reasoning.

Normally, the connection of neurons is in layers such that signals migrate from the first layer to the last layer. The first layer is the input layer while the last layer is the output layer. Contemporary ANN architectures have neural units ranging from thousands to millions and millions of connections with computing power that is equivalent to a worm brain but far simpler than a human brain. Real numbers are used to represent the signals and state of artificial neurons which are normally between 0 and 1. There is a defined limiting function (threshold function) on each unit which guides the propagation of signals – a signal must exceed the limit prior to propagating. The modification of connection weights through the use of forward stimulation is referred to as back propagation. This method is used to train the network using established correct outputs (Tahmasebi and Hezarkhani 2012). Nonetheless, success is uncertain as experience has shown that after training, some systems can adequately solve problems while others cannot. Traditionally, training has to take place in many thousands of cycles of interaction (Grefenstette et al. 2016).

The aim of ANN is to solve problems in the manner that human does. However, many neural networks are more abstract. Fresh developments in brain research often result in fresh patterns in neural networks. A recent development is the use of connections that transcend adjacent neurons to link processing layers. Researchers have intensified effort exploring new ways of manipulating signals propagated by axons. Such studies have resulted in concepts such as deep learning that pushes the limit beyond the use of a set of Boolean variables and incorporates greater complexity such as fuzzy and multi-valued variables (Grasso, Luchetta, and Manetti 2018; Mu et al. 2020). We now have newer kinds of network that are more flexible and free flowing in regards to inhibition and stimulation with links that interact in complex and chaotic manners (Kalchbrenner and Blunsom 2013). Of all the new developments, the most advanced is dynamic neural networks. They are characterized by the ability to dynamically form new neural units and connections while simultaneously disabling others using rules (Sutskever, Vinyals, and Le 2014a).

In the following section, we discuss the ANN models, learning procedures, and cost function optimization which are the bedrock for the application of neural networks in EDM in higher education domain (Cho et al. 2014).

## Neural Networks and Educational Data Mining in Higher Education

The use of technology-based pedagogies like virtual learning, e-learning, blended learning is allaying fears of researchers and stakeholders that higher education may not be able to deliver employable graduates with appropriate analytical skills and flexibility that are critical in the 21st-century knowledge economy for personal and national success (Karamouzis and Vrettos 2008; Kobayashi, Mol, and Kismihók G).

Since the new thinking, from the view point of higher education management, is data analytics, researchers have intensified efforts on using data to enhance learning and teaching by using knowledge extracted from academic records (Damaševičius 2009). An ANN-based EDM in higher education harnesses the potentials of the big data generated in higher education occasioned by growing use of mobile devices and online educational platforms (Qianyin and Bo 2015). Harnessing the humongous student-related data entails regression analysis of the observational data to identify input and output variables and the relationship between them. Next, the student data set is used to model ANN whose aim is to rigorously study the subtleness and recognize existing pattern in the student data. This involves learning by the neural network rather than programming it. The ANN continues to learn until we are confident that it has attained stability. At stability, education stakeholders are confident that the network can accurately classify students learning behaviors and predict student performance from future input data. Hence, the learning process is a metaheuristic optimization process with so many candidate solutions but with a focus on obtaining the candidate solution (network confidence) with the least cost (error) as measured by the difference between the target output of student data and expected output of the network model. This is why it is a cost function optimization problem, with the network training stopping when error is zero (i.e. no deviation between student data output and network output). This defines the level of confidence in the machine as reliable and robust tool for classifying students learning behaviors and predicting students' performance using future learner-related data (Karamouzis and Vrettos 2008). This data-based process ensures the machine gathers enough artificial intelligence about students such that it can offer reliable and robust information for informed decision making in the learning environment (Chen 2010). Such information can be used for academic advising by higher education for promoting student retention, student progression, and cost saving by critical stakeholders like course advisers, administrators, counselors, lecturers, and governments (Acevedo and Marín 2014).

Consequently, modeling, learning procedures, and cost function optimization are critical to any successful ANN-based EDM in higher education.

In eliciting information from student data, first model the data using neural network. The function (system) generated is then trained using appropriate learning procedure. Once the system is well trained as measured by attaining cost function optimization, there is sufficient confidence that the trained system can henceforth be used to classify data and predict output from future input student data.

ANN as a data-based machine learning approach uses various neurocomputing models for statistical function approximation. Common ANN models include recursive (recurrent) network, multi-layer feedforward network, single layer feedforward network, neuro-fuzzy network, deep learning network, radial basis network, extreme learning machine, among others. The aim of modeling in the context of EDM is to find a close fit for the learner-related data. The choice of a model is determined by a combination of factors.

The learning procedure used by ANN to train a machine is determined largely by the nature of data. Learning procedures include supervised learning, unsupervised learning, and reinforcement learning. While supervised learning and unsupervised learning are applied to existing data, reinforcement learning is in a scenario where data is generated as the system is being trained. This implies that learning algorithms are given data from the onset in both supervised learning and unsupervised learning while learning algorithms in reinforcement learning get their data as the system generates. However, supervised learning differs from unsupervised learning in the sense that the training data must have a target output that guides the learning process while this is absent in unsupervised learning.

Cost function optimization aims at getting an optimal solution with the least cost within defined criteria. Once this is attained, the machine is considered to have learnt enough to be relied upon for pattern recognition (classification) and data prediction. The difference between training data target output and model actual output is called error that needs to be improved upon in incremental fashion until a tolerable error estimate is attained. This is the point an optimal solution is obtained from the candidate solutions. Error estimation techniques include mean squared error (MSE), sum of squared error (SSE), and least square.

### Key Issues in Adopting Neural Networks for Educational Data Mining in Higher Education

Generally, researchers have identified six (6) broad categories of open problems that are capable of impeding EDM from delivering on its perceived benefits as frequent reports in the literature have suggested. These challenges include availability of data (Sin and Muthu 2015), accessibility of data (Slade

and Prinsloo 2013; Guenaga and Garaizar, 2016) which can be resolved through drafting of Code of Practice for EDM in organizations (Sclater 2014), interoperability (Cooper 2013), which can be addressed by adopting harmonized approaches. Others include resistance from users owing to political concerns over surveillance and personal worries over privacy (Singer 2014); impact on professional roles as benefits of EDM such as providing administrators with access to school performance data in order to facilitate critical decisions such as rewards and consequences, balanced scorecard processes, and intervention support (Seddon 2008) are being resisted by professionals, citing fear of imminent extension of accountability and managerial control and accountability; and hype as some stakeholders believe that benefits such as student retention, student progression, and cost saving may be a mirage except the challenges of EDM are tackled (Chen 2010).

### Related Work

Previous studies on neural networks and educational data mining in higher education are presented as follows. Avella et al. (Avella et al. 2016) analyzed learning analytics (LA) in higher education from the angles of methods, benefits, and challenges. The authors stressed that the learning environment and learning outcome can be enhanced using learner-related data and online educational systems. They observed that student retention, student progression, and cost saving are benefits of EDM. The study equally highlighted challenges of EDM that need to be tackled if the benefits must be realized. The study offers educational stakeholders insight into EDM's methods, benefits, and challenges so that they will be guided in applying EDM efficiently and effectively. The ultimate goal is to enhance teaching and learning in higher institutions. Nonetheless, the article did not examine ANNs for implementing EDM in higher education.

Susnea (Şusnea 2010) made efforts to advance the efficiency of ANN-based EDM in the computationally-intensive task of predicting the performance of students attending an e-learning course and classifying them.

Oancea et al. (Oancea, Dragoescu, and Ciucu 2013) identified a cardinal problem of higher education as poor results after admission, fueling early exit of students. Reasons advanced for student attrition in universities include low grades and inability to pass examinations, poor background knowledge of field of study, and lack of financial resources. The researchers posited that for university management to nip in the bud the phenomenon of high attrition rate, there is need to focus on predicting students' results so that proactive measures are taken. To practically demonstrate a possible solution, the authors used ANN in predicting students' results which were measured by their first year of study grade point average. They used a sample of 1000 students from Nicolae Titulescu University of Bucharest obtained from the last three

graduates' generations. While data set of 800 students was used for training the network, the remaining 200 students data were used for testing the network. For the experiment, the researchers used multilayer perceptron (MLP) comprised of one input layer, two hidden layers and one output layer. A variant of the resilient backpropagation algorithm was used to train the network. The input variables used include the gap between high school graduation and higher education enrolling, student age, and GPA at high school graduation, which constitute the students profile at the time of enrolling at the university. The authors concluded that a well-trained network will predict students' results and offer invaluable academic advise to management on ways to circumvent students' dropout. The study drew attention to the potentials of ANN for academic advising even though it is not an SLR and fell short to highlighting open issues of ANN-based EDM.

Papamitsiou and Economides (Papamitsiou and Economides 2014) relied on practical evidence from empirical research to offer current knowledge on EDM and its impact on adaptive learning. The study stressed key objectives of the adoption of EDM in generic educational strategic planning. Using an SLR approach, the researchers x-rayed the literature on experimental case studies conducted in EDM between 2008 and 2013. With the aid of search items, 209 pieces of research work were identified even though only 40 key studies were used after applying inclusion criteria. Thereafter, the published papers were categorized after analyzing their methodology, research questions, and findings. The harvested studies were evaluated using non-statistical methods and findings interpreted. The results showed four major directions of empirical research in EDM. Based on findings, the authors pointed out the value-add of EDM in educational strategic planning and stressed the importance of further implications. From pedagogical and technical considerations, they outlined key questions for further investigation. Though the article is an SLR that offered current knowledge on EDM using empirical experimental case studies from 2008 to 2013, this present study is more current using research studies from 2010 to 2018. Also, the current study specifically examines ANN for EDM in higher education unlike (Zissis, Xidias, and Lekkas 2015) that focused on the rather generic EDM application in educational strategic planning.

Borkar and K. Rajeswari (Borkar and Rajeswari 2014) emphatically stated that EDM is a promising discipline that impacts on the education sector. To buttress their stance, the authors experimented with student data, applying ANNs as a data mining technique to evaluate students' performance. Select student attributes were used to elicit rules using association rule mining while ANN was deployed to check accuracy of the results and predict student performance. In any case, the focus of the study was not on using SLR to identify gaps and challenges of ANN-based EDM in higher education, let alone providing direction for future research in the nascent but promising field of EDM.

Siri (Siri 2015) researched the use of ANN model to offer academic advice for the design of educational interventions to help students who are at the risk of dropping out. The work was part of efforts to contribute to ongoing debates on how to use data mining techniques such as ANNs to mitigate student failure and scale up educational processes. For experiment, the empirical study which is an integral part of an existing project that investigates the initial stage of student transition to university, used a population of 810 students. The students were enrolled in the 2008–2009 academic year for the first time at the University of Genoa in a health care professions degree course. The student data were obtained from primary sources such as administrative data related to student careers, telephone interviews with students, and ad hoc survey. The outcome of the study after analyzing data and information using ANN indicated that the ANN accurately predicted percentages of the students that belonged to different groups, including 76% of the cases (students) belonging to the category of dropouts. In any case, the study did not examine EDM from the viewpoint of modeling, learning procedures, and cost function optimization for the purpose of unveiling gaps and challenges of the ANN-based EDM for future research direction, which is the aim of the present study.

Yorek and Ugulu (Yorek and Ugulu 2015) emphasized that ANN is a new approach applied in educational qualitative data analysis to ascertain characteristics of students during learning. They opined that the ANN model could be trained to elicit qualitative results from humungous amount of learner-related categorical data. To test their assertion, a cascade-forward back-propagation neural network (CFBPN) model was developed and used to analyze quantitative student data in a bid to ascertain their attitudes. The researchers collected the data using a conceptual understanding test that included open-ended questions. The outcome of the study indicated that the use CFBFN model to analyze educational research that examines learning attitudes, behaviors, or beliefs helps in obtaining detailed information about data analyzed and by extension subtle information about the characteristics of the subjects (participants). The study gave credence to the assertion that ANN is at the cutting edge of science and technology though it did not dwell on the challenges that had hindered its less than expected usage in the education sector.

Bernard et al. (Bernard et al. 2015) used ANN to identify learning styles with a view to customizing learning for high performance, learning satisfaction, and reducing the time required for learning. They opined that adaptive learning systems offer personalized content to students taking into cognizance their learning styles. Though questionnaires can be used to identify students' learning styles, the authors hinted that the approach has several demerits. To overcome these challenges, research has been carried out on automatic approaches that can identify learning styles. Since this line of research is still at infancy coupled with the fact that current approaches need significant

improvement before their effective use in adaptive systems, the authors succumbed to using ANN to identify students' learning styles. The study evaluated the ANN approach using data from 75 students and it was discovered that it outperformed other approaches in terms of accuracy of identify learning styles. With accurate learning style identification, quality academic advice could be offered students by way of adaptive systems or by informed teachers who know precisely their students' learning styles. The authors concluded that such informed academic advising leads to greater learning satisfaction, higher performance, and reduced learning time. Despite focusing on enriching learning experience using ANN, the focus was not primarily to identify gaps and challenges confronting the growth of ANN-based EDM. Also, the study is not an SLR.

Pavlin-Bernardić et al. (Pavlin-Bernardić, Ravić, and Matić 2016) wondered why the less than proportionate usage of ANN in educational psychology despite its wide-ranging use in the prediction and classification of different variables. To practically demonstrate the potentials of ANN in the field of education psychology in particular, and higher education in general, the study examined the accuracy of ANN in predicting students' general giftedness. A sample size of 221 participants who were fourth grade students from a Croatian elementary school was used. The input variables used were school grades, nominations by teachers and peers, parents' education, and earlier school readiness assessment. The students were classified as gifted or nongifted based on the output variable which was the result on the Standard Progressive Matrices. The study demonstrated the potentials of ANN in educational psychology and by extension higher education, and therefore should be further explored. The study buttressed the motivation for the present study which asserts that ANN-based EDM in higher education has received little attention in the literature despite its huge potentials for transforming higher education.

Bahadir (Bahadır 2016) gave credence to the assertion that ability to predict student performance is an invaluable asset for academic advisory. The work which focused mainly on predicting academic success in graduate programmes, reiterated that the ability to predict success of students in a graduate programme is useful to educational institutions for the development of strategic programmes that improve student performance during their stay in the institution. The author presented the outcomes of an experimental comparison study between Logic Regression Analysis (LRA) and ANN for the prediction of academic success of mathematics teachers when they enter graduate education. The model was trained and tested with a sample size of 372 student profiles. The researcher measured the strength of the model through LRA and found out that the mean correct success rate of students for ANN was higher than LRA. Specifically, it was reported that the successful prediction rate of the back-propagation neural network was 93.02% as

compared 90.75% of LRA. The study underpins the relevance of ANN in higher education as a mechanism for informed academic advisory even though it did not discuss the challenges of an ANN-based EDM let alone suggesting ways of improving its usage for higher education.

Okewu and Daramola (Okewu and Daramola 2017) reported on the design of a multi-tier layered architecture as an EDM system for a consortium of Nigeria universities. Their aim was to outline ways of tackling technical challenges confronting EDM implementation which could hinder the realization of benefits it offers higher education. The technical challenges include interoperability, scalability, security, amongst others. The proposed EDM framework which doubles as an academic advisory system has inbuilt mechanisms that address these open problems. The researcher opined that as a tool for academic advising, the framework can offer advice at many layers of granularity like department, individual, university, and national with a view to facilitating student progression, student retention, and cost saving. Aside tackling open issues of EDM in higher education, the study offered empirical steps for the application of EDM in an African context which the authors claimed has received little attention in the literature. Despite addressing technical challenges of EDM, the work was not an SLR. Neither did it approach EDM in higher education from the ANN view point.

In summary, none of the previous SLRs examined ANN-based EDM in higher education from the view point of modeling, learning algorithms, and cost function optimization in a bid to forge understanding of EDM from these key aspects. In this paper, we attempt to promote robust understanding of the EDM in higher education landscape in terms of research and practice from the perspective of ANNs. Also, exploring the accuracy potentials of neural networks in EDM will ensure visible quality educational service delivery. The observation by education stakeholders that EDM is presently a mere hype without strong evidence of concrete impact on higher education (Ali et al. 2013),(Macfadyen and Dawson 2012) is concerning and can adversely affect the future of EDM. Hence, this ANN-based study is crucial for both researchers and stakeholders of EDM in higher education as it highlights how ANN can reposition EDM.

## Research Method

An SLR involving the use neural networks for building EDM systems in higher education is an intelligent software engineering initiative. We carried out this systematic study using established guidelines for performing SLRs in Software Engineering (Kitchenham, Charters, and Kuzniarz 2015) while Host and Orucevic-Alagic (Höst and Oručević-Alagić 2011) provided the steps for conducting the SLR.

### *Data Sources – Academic Databases*

We started by defining adequate scope for the research, spanning ANN modeling, learning procedures, cost function optimization, and challenges. Thereafter, we generated research questions as well as identified keywords. The search for the keywords was done in top-range and quality academic databases such as Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and Springerlink.

In a bid to elicit plenty important papers, we extended the list of databases. After harvesting resourceful primary studies, we generated a validation list in order to evaluate search strings and confirm the adequacy of the searches.

The research studies were extracted from the following academic databases in the course of the SLR.

1)IEEEXplore (ieeexplore.ieee.org)
2)ACM DL portal (dl.acm.org)

### 3) Springer Link (www.link.springer.com*)*

3)Scopus (www.scopus.org)
4)Web of Science (www.webofknowledge.com)

### *Data Retrieval*

To cover a wide range of publications on neural networks and EDM in higher education, we tried series of combinations of keywords. The tactics equally aided the verification of synonyms as used in the literature. The search process involved leveraging on the Boolean operators OR and AND for combining these search items. The format of the search process is as follows:

*("Smart Education" OR*
*"Smart Learning" OR*
*"Academic Advising" OR*
*"Educational data mining" OR*
*"Learning analytics" OR*
*"Academic analytics" OR*
*"Intelligent tutoring") AND*
*("Artificial Neural Network" OR*
*"Neural Networks" OR*
*"Machine Learning")*

In Scopus, we limited the search results to Journals and Conference Proceedings to omit Editorials and Proceeding Volumes and Books.

Nonetheless, we observed that even with this technique, majority of the displayed papers had been extracted in previous individual searches. Yet another initiative was placing requests to authors via ResearchGate for papers that we

**Table 1.** data retrieval.

| SN | Academic Database | No. of Research Studies (Papers) Extracted |
|----|------------------|-------------------------------------------|
| 1. | ACM Digital Library | 138 |
| 2. | IEEE Xplore Library | 205 |
| 3. | Scopus | 395 |
| 4. | Web of Science | 296 |
| 5. | Springerlink | 53 |
|    | Total | 577 |

could not extract in full-text form but were viewed relevant. During the search and retrieval process, some articles harvested in one database were found in other databases. We concluded, based on this finding that a strong relationship exists between academic databases with respect to article content. To forestall possible errors from the overlapping effect and their potential consequences for the outcome of our study, we excluded papers seen in one academic database from the count of others. We observed during the extraction process that many papers in Web of Science were also housed in Scopus, apparently the largest academic database. In the same vein, some papers harvested from Springerlink, ACM Digital Library, and IEEE Xplore were present in Scopus. Following painstaking and meticulous efforts at ensuring that each paper was uniquely counted, the data retrieval process produced an outcome as shown in Table 1.

We then applied Inclusion and Exclusion Criteria to the 577 articles to elicit only important papers. The inclusion criteria are:

i)The topic, abstract, or body of article must make reference to neural networks or educational data mining in higher education.

ii)The most current publication is selected in the event two or more publications report same research.

The exclusion criteria are:

i)Exclusion of publications that discussed issues other than neural networks and educational data mining in higher education.

ii)Materials without scientific or academic references were exempted even if they had professional perception.

## Findings of the Study

In the literature, the three priority concerns of researchers in ANN modeling, learning algorithms, and cost function optimization are accuracy (Deng et al. 2012), processing speed (Zissis, Xidias, and Lekkas 2015) and memory space utilization (Ruslan and Joshua 1958–71). The findings of the study based on four research questions are presented below.

## *Rq 1*

ANN is a data-based machine learning technique while educational data mining is a data-assisted technology-enhanced learning. Action researches and advances in modeling are aimed at producing networks with better generalization performance and learning speed. Applying ANN to learner-related data implies that the statistical functions inherent in the data have to be adequately mapped. The ability to map data in a close-fit fashion using different neural networks models gained researchers attention. A possible explanation for this trend is that the two main parameters for measuring the success of ANN learning are ability to approximate the functions (Mnih et al. 2015) in observational data (function approximation or regression) and ability to predict accurately (Socher and Lin 2011). The literature also confirmed that modeling provides bedrock for representation learning which in turn enhances predictive mining and pattern recognition (classification). In developing models, research efforts are geared toward efficiency and effectiveness. The three main categories of neural network models are recurrent network, multi-layer feedforward network, and single layer feedforward network (Dahl et al. 2012). Modeling with recurrent network got the highest attention; this is followed by multi-layer feedforward network: while the least number of publications focused on single layer feedforward network. Specific data on the key topics of ANN are presented in Table 3, while the distribution of key topics as disseminated in research publication outlets is shown in Figure 1.

Vast majority of researchers (47.2%) that focused on the models commented on the recurrent (recursive) network and its capability to map data in a close-fit format. A total of 77 papers which represent investigated recurrent network as shown in Table 3. They stressed the need for an end-to-end
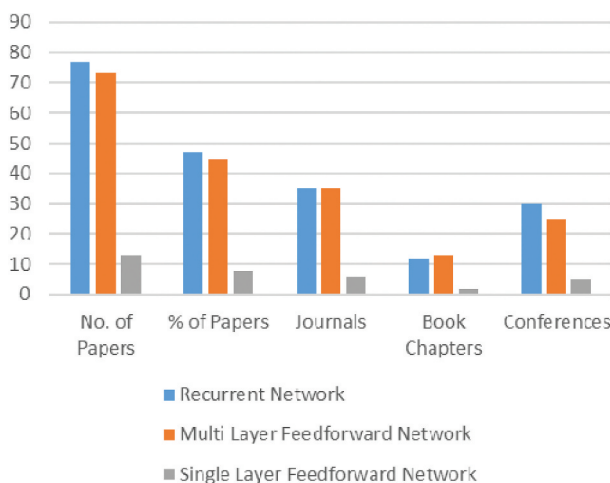


**Figure 1.** Overview of distribution of topics on Neural Network Models in publication outlets.

approach to regression (function approximation) so that vital statistical functions/features of training data such as online data that covers academic and administrative units (Komenda et al. 2015) are adequately represented. There is a consensus that recurrent networks like backpropagation models receive strong attention, in part owing to their ability ensure thorough search for an optimal solution with the least cost. This they achieve by recursively adjusting assigned weights and recalculating error estimates until the desired actual output from the neural network is attained. This actual model output, which is closest to the training data target output provides confidence that the ANN has been sufficiently trained for purposes of being used for generalization and predicting output from future student input data. This way, an ANN-based EDM can extract vital information from student data that will aid educational institutions in the aspects of scaling up student retention, improving student success, and reducing the burden of accountability (Dietz-Uhler and Hurn 2013). Examples of recurrent network are Boltzmann machine (Courville, Bergstra, and Bengio 2011) and recursive deep neural network (Hutchinson, Deng, and Yu 2012).

The second key concern of authors in the modeling of tasks is the use of multi-layer feedforward network for adequate mapping of the statistical functions of sample data. In this network, there are many hidden layers between the input and the output (Schmidhuber 2015a). Hence, multiple-processing layers are used by the training algorithm and the layers are made up of many linear and non-linear transformations. An example is deep learning or deep neural network. There were 73 research studies that concentrated on this area, representing 44.8% of the total number of papers on neural network modeling. The researchers posited that multiple layers in a feedforward network particularly in the hidden layer where the bulk of neural computations takes place will guarantee sufficient training of machine so as to boost confidence of users in its outcomes. Nonetheless, they pointed out that though such rigorous and multi-layer training is good for complex and mission-critical applications (task areas), it comes at a cost which includes processing time and memory storage. This model is applicable in fields like automatic speech recognition, computer vision, audio recognition, bioinformatics, and natural language processing. The results produced are comparable to human experts, and in some instances, superior (Okewu and Daramola 2017).

The least researched topic in the context of ANN modeling is single layer feedforward network and an example is Extreme Learning Machine (ELM). Thirteen (13) papers accounting for 8% of research commitments on ANN modeling evaluated the potentials of single feedforward networks. Single feedforward network is only adequate for simple tasks whose learning procedures are simplistic. However, when applied to complex tasks, it inadequacies manifest greatly. This is because the machine learning mapping of the statistical function mathematically expressed as F: x → y is not rigorous enough to

capture all the parameters of observational data and perform comprehensive neural computations in a single layer of hidden neurons. Such parameters include input data, assigned weights, introduced constant, among others. Single layer network is useful for classification and regression problems (Damaševičius 2009).

### Rq 2

The aim of applying ANN to learning-related data in in higher education is to harness its learning potentials for training educational data mining systems. Hence, the introduction of ANN in the education sector is to encourage data-based machine learning (Courville, Bergstra, and Bengio 2011) rather than rule-based machine learning (Courville, Bergstra, and Bengio 2012). It is believed that ANN-based EDM will solve the problem of monitoring and regulating behaviors of learners in a virtual learning environment. Such monitoring cannot be done physically as in traditional face-to-face pedagogy. Training a machine is a learning procedure that involves tasks and learning algorithms.

Learning procedures were given good attention in the literature going by the number of research studies on the subject that was found. Most of the research studies that discussed educational data mining focused on EDM implementation in universities though some studies examined the subject in other segments of higher education. Majority of implementations were in higher institutions in Europe and North America. Overall, the research studies examined three categories of machine learning – supervised learning, unsupervised learning, and reinforcement learning.

The data on key topics on learning procedure is shown in Table 4, while the distribution of key topics in research publication outlets is shown in Figure 2.

As indicated in Table 4 and Figure 2, most authors (49.7%) who discussed learning procedures focused on supervised learning. They argue that supervised learning is good for in computationally intensive application areas like educational data mining where there is a clear-cut input-output pair; the output variable serves as the teacher given its labeled classification. Students' observational data normally appear in the input-output pair format where the input component could be a multi-variate vector while the output is a bi-labeled or multi-labeled translating into binary classification or multi-labeled variable. A classic example is relationship mining between scores in courses taken by a student and class of degree obtained. This is a classification problem as the values of the output (class of degree) are discrete (First Class, Second Class, Third Class, Pass). Hence, tasks like pattern recognition and predictive mining are handled using supervised learning in domains like education, finance, engineering, medicine, just to mention a few. In supervised learning, learning algorithms are guided by a training data target output to work toward
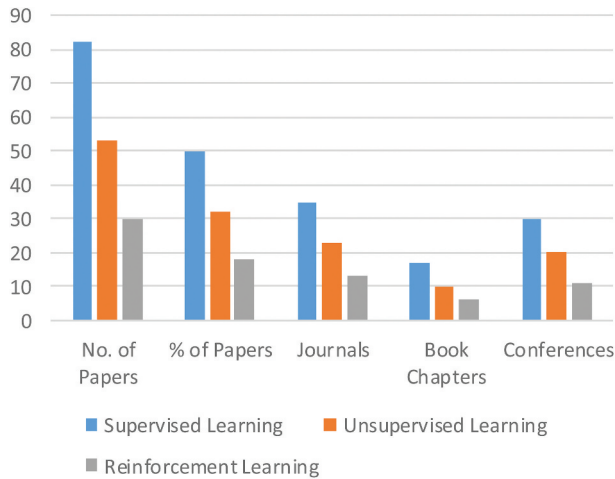
**Figure 2.** Distribution of topics on Learning Procedure publication outlets.

and they usually adopt metaheuristic search techniques to obtain an optimal solution with least cost function.

Unsupervised learning received moderate attention. We have 53 papers in this category, constituting 32.1% of primary studies on learning procedures. The authors advocated that a task and application area where there is no clear-cut input-output dichotomy but rather the emphasis to determine relationship pattern between the input variables should be subjected to unsupervised learning as the befitting learning procedure. The corroborated that since such statistical functions don't have a target output that could serve as the teacher, the neural network should be trained by understanding the hidden and interesting pattern between the input variables. Tasks that typically use unsupervised learning (Hutchinson, Deng, and Yu 2012) include estimation problems and the applications include the applications include filtering, clustering, compression and estimation of statistical distributions. In unsupervised learning, the learning algorithms do not have a target to work toward.

The least researched learning procedure is reinforcement learning with 30 papers, constituting 18.2% of total research on learning procedures. These primary studies emphasized that unlike supervised learning and unsupervised learning where sample data exists for modeling, such does not exist in reinforcement learning. Rather the tasks and application areas are such that data are system-generated. As the system is operational, data is continuously generated and ANN leverages on the data to understand data pattern and get well trained to mimic hidden behaviors of the data. Therefore, the learning algorithm used by the ANN in reinforcement learning procedure should be adapted to system-generated data. Examples of tasks and application areas that use reinforcement learning are included.

### RQ3

Attempts have been made to develop cost function optimization methods (Deng, Yu, and Platt 2012) that will offer optimal solution. The training process of applying neural networks to learner-related data entails that the outcome of the neural network referred to as actual output should be close to the target output of the observational data within acceptable limits for the training to be considered successful. It is only when this feat is attained that the ANN is considered stable and suitable for classifying and predicting output from future input data (Mukherjee et al. 2006). The variance between target output of observational data and actual output of the model (ANN) is considered a cost that must be minimized. Hence, it is called cost function optimization and the acceptable solution must have the least cost of all the candidate solutions. Cost function optimization is a multi-step search approach and requires metaheuristic algorithms like genetic algorithm, tabu search, particle swarm optimization, ant colony optimization, among others (Socher and Lin 2011).

From Table 2, it can be inferred that a total of 103 papers touched on issues of cost function optimization evaluation (MLCO+MLC+C) out of the 190 papers used as primary studies, which represents 54.2%. Overtime, attempts have been made to develop cost function optimization methods (Fan et al. 2015) that will offer optimal solution. The training process of applying neural networks to learner-related data entails that the outcome of the neural network referred to as actual output should be close to the target output of the observational data within acceptable limits for the training to be considered successful. It is only when this feat is attained that the ANN is considered suitable for classifying and predicting output from future input data. The variance between target output of observational data and actual output of the model (ANN) is considered a cost that must be minimized. Hence, it is called cost function optimization and the acceptable solution must have the least cost of all the candidate solutions. Cost function optimization is a multi-step search approach and requires metaheuristic algorithms like genetic algorithm, tabu search, particle swarm optimization, ant colony optimization, among others.

The topic map for the three cost function optimization methods is presented in Table 5 while the distribution of key topics in publication outlets is graphically illustrated in Figure 3.

From the topic shown in Table 5 and Figure 3, it is obvious that 50.5% of the researchers that investigated cost function optimization focused on mean squared error (MSE) otherwise referred to as mean squared deviation. They argued that finding the mean square of the error (variance) between an observed value and the fitted value computed by a model will offer the best fit in the mean squared sense. The studies pointed out that the most important

**Table 2.** Topic Map FOR NEURAL NETWORK MODELING.

| SN | Topic | List of papers |
|---|---|---|
| 1. | Recurrent Network | **Total = 77 (47.2%); Journals = 35; Book Chapters = 12; Conferences = 30** <br> (Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Schmidhuber 2015a), (Tahmasebi and Hezarkhani 2012), (Grefenstette et al. 2016), (Mu et al. 2020), (Grasso, Luchetta, and Manetti 2018), (Kalchbrenner and Blunsom 2013), (Sutskever, Vinyals, and Le 2014a), (Cho et al. 2014), (Şusnea 2010), (Papamitsiou and Economides 2014), (Siri 2015), (Pavlin-Bernardić, Ravić, and Matić 2016), (Bahadır 2016), (Deng et al. 2012), (Ruslan and Joshua 1958–71), (Socher and Lin 2011), (Dahl et al. 2012), (Mohamed, Dahl, and Hinton 2012), (Komenda et al. 2015), (Courville, Bergstra, and Bengio 2011), (Hutchinson, Deng, and Yu 2012), (Deng, Yu, and Platt 2012), (Deng and Yu 2011), (Vincent et al. 2010), (Hinton and Salakhutdinov 2012), (Graupe 2016), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (Sak, Senior, and Beaufays 2014a), (Zen and Hasim 2015), (Fan et al. 2015), (Forouzanfar et al. 2014), (Ojha, Abraham, and Snášel 2017), (Cireşan et al. 2011a), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Hinton et al. 2012a), (Cireşan et al. 2010a), (Martines, Bengio, and Yannakakis 2013), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Zhu, Miao, and Qing 2014), (Ciresan et al. 2013), (Zhong, Liu, and Liu 2011), (Bengio et al. 2003), (Socher 2013), (Huang et al. 2013), (Mesnil et al. 2015), (Gao et al. 2014), (Qu et al. 2018), (Wallach, Dzamba, and Heifets 2015), (Tkachenko 2015), (Elkahky, Song, and He 2015), (Choi et al. 2016), (Bengio, Courville, and Vincent 2013), (Sak, Senior, and Beaufays 2014b), (Li and Wu 2015), (Zen and Sak 2015), (Gers and Schmidhuber 2001), (Mikolov et al. 2010), (Bengio, Boulanger-Lewandowski, and Pascanu 2013), (Campagni et al. 2015), (Ferreira and Andrade 2014), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |
| 2. | Multi Layer Feedforward Network | **Total = 73 (44.8%); Journals = 35; Book Chapters = 13; Conferences = 25** <br> (Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Schmidhuber 2015a), (Tahmasebi and Hezarkhani 2012), (Grefenstette et al. 2016), (Mu et al. 2020), (Grasso, Luchetta, and Manetti 2018), (Sutskever, Vinyals, and Le 2014a), (Şusnea 2010), (Papamitsiou and Economides 2014), (Siri 2015), (Pavlin-Bernardić, Ravić, and Matić 2016), (Bahadır 2016), (Deng et al. 2012), (Ruslan and Joshua 1958–71), (Dahl et al. 2012), (Mohamed, Dahl, and Hinton 2012), (Komenda et al. 2015), (Courville, Bergstra, and Bengio 2011), (Hutchinson, Deng, and Yu 2012), (Deng, Yu, and Platt 2012), (Deng and Yu 2011), (Vincent et al. 2010), (Hinton and Salakhutdinov 2012), (Graupe 2016), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (Fan et al. 2015), (Forouzanfar et al. 2014), (Ojha, Abraham, and Snášel 2017), (Cireşan et al. 2011a), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Hinton et al. 2012a), (Widrow et al. 2013), (Cireşan et al. 2010a), (Martines, Bengio, and Yannakakis 2013), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Zhu, Miao, and Qing 2014), (Abdel-Hamid et al. 2014), (Ciresan et al. 2013), (Zhong, Liu, and Liu 2011), (Bengio et al. 2003), (Socher 2013), (Huang et al. 2013), (Gao et al. 2014), (Qu et al. 2018), (Wallach, Dzamba, and Heifets 2015), (Elkahky, Song, and He 2015), (Chicco, Sadowski, and Baldi 2014), (Silver et al. 2016), (Bengio, Courville, and Vincent 2013), (Deng 2014), (Hinton et al. 2012b), (Deng, Hinton, and Kingsbury 2013), (Deng et al. 2013), (Deng, Abdel-Hamid, and Yu 2013), (Sainath et al. 2013), (Szegedy and Erhan 2013), (Mikolov et al. 2010), (Fischer and Igel 2014), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |
| 3. | Single Layer Feedforward Network | **Total = 13 (8%); Journals = 6; Book Chapters = 2; Conferences = 5** <br> (Forouzanfar et al. 2014), (Ojha, Abraham, and Snášel 2017), (Zhu et al. 2015), (Zhu, Miao, and Qing 2014), (Bengio et al. 2003), (Gao et al. 2014), (Bengio 2012) (Coates, Lee, and Ng 2011), (Campagni et al. 2015), (Ferreira and Andrade 2014), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |

application of MSE just like other error estimation methods such as sum of squared error and least squares is in data fitting. The data fitting application ensures that the outcome of an ANN model approximates the statistical function of observational data with minimal acceptable cost. Hence, it is a cost function optimization initiative.

**Table 3.** Topic Map for Learning Procedures.

| SN | Topic | List of papers |
|---|---|---|
| 1. | Supervised Learning | **Total = 82 (49.7%); Journals = 35; Book Chapters = 17; Conferences = 30**<br>(Poggio and Rosasco et al. 2012), (Sidhu and Caffo 2014), (Mukherjee et al. 2006), (Şusnea 2010), (Siri 2015), (Pavlin-Bernardić, Ravić, and Matić 2016), (Bahadır 2016), (Deng et al. 2012), (Mnih et al. 2015), (Courville, Bergstra, and Bengio 2011), (Courville, Bergstra, and Bengio 2012), (Hutchinson, Deng, and Yu 2012), (Hinton and Salakhutdinov 2012), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (Cireşan et al. 2011a), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Ciresan, Meier, and Schmidhuber 2012a), (Ciresan et al. 2011a), (Krizhevsky, Sutskever, and Hinton 2012), (Hinton et al. 2012a), (Widrow et al. 2013), (Schmidhuber 2015c), (Cireşan et al. 2010a), (Martines, Bengio, and Yannakakis 2013), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Zhu et al. 2015), (Zhu, Miao, and Qing 2014), (Abdel-Hamid et al. 2014), (Yu and Deng 2010), (Ciresan et al. 2013), (Zhong, Liu, and Liu 2011), (Bengio et al. 2003), (Socher 2013), (Shen et al. 2014), (Huang et al. 2013), (Mesnil et al. 2015), (Gao et al. 2014), (Qu et al. 2018), (Tkachenko 2015), (Elkahky, Song, and He 2015), (Chicco, Sadowski, and Baldi 2014), (Choi et al. 2016), (Silver et al. 2016), (Bengio, Courville, and Vincent 2013), (Krizhevsky, Sutskever, and Hinton 2012), (Deng 2014), (Hinton et al. 2012b), (Deng, Hinton, and Kingsbury 2013), (Deng et al. 2013), (Sak, Senior, and Beaufays 2014b), (Li and Wu 2015), (Zen and Sak 2015), (Deng, Abdel-Hamid, and Yu 2013), (Sainath et al. 2013), (Yu and Deng 2014), (Szegedy and Erhan 2013), (Gers and Schmidhuber 2001), (Sutskever, Vinyals, and Le 2014b), (Mikolov et al. 2010), (Bengio, Boulanger-Lewandowski, and Pascanu 2013), (Schmidhuber 2015d), (Unterthiner et al. 2015), (Fischer and Igel 2014), (Coates, Lee, and Ng 2011), (Hastie, Tibshirani, and Friedman 2009), (Mohri, Rostamizadeh, and Talwalkar 2012c), (Budgaga et al. 2016), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |
| 2. | Unsupervised Learning | **Total = 53 (32.1%); Journals = 23; Book Chapters = 10; Conferences = 20**<br>(Sidhu and Caffo 2014), (Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Şusnea 2010), (Mnih et al. 2015), (Courville, Bergstra, and Bengio 2011), (Courville, Bergstra, and Bengio 2012), (Hutchinson, Deng, and Yu 2012), (Hinton and Salakhutdinov 2012), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Zhu et al. 2015), (Yu and Deng 2010), (Bengio et al. 2003), (Socher 2013), (Shen et al. 2014), (Huang et al. 2013), (Mesnil et al. 2015), (Gao et al. 2014), (Qu et al. 2018), (Wallach, Dzamba, and Heifets 2015), (Tkachenko 2015), (Elkahky, Song, and He 2015), (Hinton et al. 2012b), (Deng, Hinton, and Kingsbury 2013), (Deng et al. 2013), (Sak, Senior, and Beaufays 2014b), (Li and Wu 2015), (Zen and Sak 2015), (Deng, Abdel-Hamid, and Yu 2013), (Sainath et al. 2013), (Yu and Deng 2014), (Szegedy and Erhan 2013), (Gers and Schmidhuber 2001), (Sutskever, Vinyals, and Le 2014b), (Mikolov et al. 2010), (Bengio, Boulanger-Lewandowski, and Pascanu 2013), (Schmidhuber 2015d), (Unterthiner et al. 2015), (Fischer and Igel 2014), (Coates, Lee, and Ng 2011), (Hastie, Tibshirani, and Friedman 2009), (Mohri, Rostamizadeh, and Talwalkar 2012c), (Budgaga et al. 2016), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |
| 3. | Reinforcement Learning | **Total = 30 (18.2%); Journals = 13; Book Chapters = 6; Conferences = 11**<br>(Sidhu and Caffo 2014), (Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Şusnea 2010), (Mnih et al. 2015), (Courville, Bergstra, and Bengio 2011), (Courville, Bergstra, and Bengio 2012), (Hutchinson, Deng, and Yu 2012), (Hinton and Salakhutdinov 2012), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Zhu et al. 2015), (Yu and Deng 2010), (Bengio et al. 2003), (Yu and Deng 2014), (Gers and Schmidhuber 2001), (Sutskever, Vinyals, and Le 2014b), (Schmidhuber 2015d), (Unterthiner et al. 2015), (Fischer and Igel 2014), (Coates, Lee, and Ng 2011), (Hastie, Tibshirani, and Friedman 2009), (Mohri, Rostamizadeh, and Talwalkar 2012c), (Mohri, Rostamizadeh, and Talwalkar 2012c), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |

Second in the pecking order of cost function optimization evaluation is the sum of squared error (SSE). 36.8% of authors are of the view that effective and efficient data fitting between actual output of ANN model and target output of training data can be achieved use SSE. The process involves iterative

**Table 4.** Topic Map for Cost Function Optimization Methods.

| SN | Topic | List of papers |
|---|---|---|
| 1. | Mean Squared Error (MSE) | **Total = 52 (50.5%); Journals = 26; Book Chapters = 8; Conferences = 18**<br>(Bengio et al. 2003), (Socher 2013), (Shen et al. 2014), (Huang et al. 2013), (Mesnil et al. 2015), (Gao et al. 2014), (Qu et al. 2018), (Wallach, Dzamba, and Heifets 2015), (Tkachenko 2015), (Elkahky, Song, and He 2015), (Chicco, Sadowski, and Baldi 2014), (Choi et al. 2016), (Silver et al. 2016), (Bengio, Courville, and Vincent 2013), (Krizhevsky, Sutskever, and Hinton 2012), (Deng 2014), (Hinton et al. 2012b), (Deng, Hinton, and Kingsbury 2013), (Deng et al. 2013), (Sak, Senior, and Beaufays 2014b), (Li and Wu 2015), (Zen and Sak 2015), (Deng, Abdel-Hamid, and Yu 2013), (Sainath et al. 2013), (Szegedy and Erhan 2013), (Gers and Schmidhuber 2001), (Mikolov et al. 2010), (Bengio, Boulanger-Lewandowski, and Pascanu 2013), (Coates, Lee, and Ng 2011), (Campagni et al. 2015), (Ferreira and Andrade 2014), (Palmer 2013), (Retalis et al. 2006), (Shum and Deakin Crick 2012) |
| 2. | Sum of Squared Errror (SSE) | **Total = 38 (36.8%); Journals = 16; Book Chapters = 8; Conferences = 14**<br>(Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Schmidhuber 2015a), (Tahmasebi and Hezarkhani 2012), (Grefenstette et al. 2016), (Mu et al. 2020), (Grasso, Luchetta, and Manetti 2018), (Kalchbrenner and Blunsom 2013), (Sutskever, Vinyals, and Le 2014a), (Cho et al. 2014), (Şusnea 2010), (Papamitsiou and Economides 2014), (Siri 2015), (Pavlin-Bernardić, Ravić, and Matić 2016), (Bahadır 2016), (Deng et al. 2012), (Ruslan and Joshua 1958–71), (Socher and Lin 2011), (Dahl et al. 2012), (Mohamed, Dahl, and Hinton 2012), (Komenda et al. 2015), (Courville, Bergstra, and Bengio 2011), (Hutchinson, Deng, and Yu 2012), (Deng, Yu, and Platt 2012), (Deng and Yu 2011), (Vincent et al. 2010), (Hinton and Salakhutdinov 2012), (2016), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (Sak, Senior, and Beaufays 2014a), (Cireşan et al. 2010a), (Martines, Bengio, and Yannakakis 2013), (Ciresan, Meier, and Schmidhuber 2012b), (Bassel et al. 2011), (Lu, Plataniotis, and Venetsanopoulos) |
| 3. | Least Squares | **Total = 13 (12.6%); Journals = 7; Book Chapters = 2; Conferences = 4**<br>(Poggio and Rosasco et al. 2012), (Mukherjee et al. 2006), (Şusnea 2010), (Komenda et al. 2015), (Courville, Bergstra, and Bengio 2011), (Hinton and Salakhutdinov 2012), (Waxman, Graupe, and Carley 2010), (Graupe 2013), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Retalis et al. 2006) |

**Table 5.** Topic Map For Open Issues Of Neural Network-Based Edm For He

| SN | Topic | List of papers |
|---|---|---|
| 1. | Hardware Issues | **Total = 8 (34.7%); Journals = 4; Book Chapters = 2; Conferences = 2**<br>(Hutchinson, Deng, and Yu 2012), (D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen 2012), (Yuan, Ho, and Lin 2012), (Langley 2011b), (Kumar and Kumar 2012), (Li and Wu 2015), (Zen and Sak 2015), (Sze et al. 2017) |
| 2. | Training Issues | **Total = 6 (26.1%); Journals = 3; Book Chapters = 1; Conference = 2**<br>(Lu, Plataniotis, and Venetsanopoulos), (Yu and Deng 2010), (Nguyen, Yosinski, and Clune 2014), (Hinton et al. 2012), (Bengio 2012), (Sze et al. 2017), (Hinton 2010), (Fischer and Igel 2014) |
| 3. | Theoretical Issues | **Total = 6 (26.1%); Journals = 3; Book Chapters = 1; Conference = 2**<br>(Hutter 2012), (Oudeyer 2010), (Rajani 2011), (Gopnik 2017), (Marcus 2017), (Ciresan et al. 2013) |
| 4. | Quality Issues | **Total = 3 (13.0%); Journals = 2; Book Chapters = 0; Conferences = 1**<br>(Drachsler and Greller 2016), (Hoel and Chen 2014), (Palmer 2013) |

refinement of the residue (error estimate) using SSE until an error estimate that is within acceptable limit of tolerance is attained. At this point, the output from the ANN function is considered optimal solution with the least cost out of all the candidate solutions in the solution space transversed through by the learning machine. It is equally believed that at this point, the ANN machine has been adequately trained for performing tasks such as pattern recognition, function approximation, and predictive mining/analytics.

**Figure 3.** Distribution of key topics on Cost Function Optimization methods in publication outlets.

Other researchers focused on fitting a set of data points using least-squares approximation, though they were in the least. As portrayed in Table 5 and Figure 3, just 12.6% of authors reported that reported on cost function optimization were interest in least squares method. In any case, their findings are pretty valuable. They asserted that the best fit in function approximation of sample data using ANN in the context of least-squares minimizes the sum of squared residuals. Residual refers to the variance between observed value and fitted value produced by the ANN model). The residual value obtained at each step of the iterative refinement is improved upon by adjusting assigned weights and recalculating weighted output and the process so continue until an optimal solution with least cost is obtained within specified criteria. Though training network this way takes time and memory space, it ensures delivery resilience in terms of pattern recognition (classification), regression (function approximation) and predictive mining (analytics) of future data.

### RQ4

In the literature, there are worries that except the open problems of neural networks are addressed, their perceived benefits of handling large input parameters, accurate prediction, and reliable pattern recognition may be a mere hype (Deng and Yu 2011). By extension, except the challenges confronting neural network-based EDM in higher education are tackled, achieving EDM's perceived benefits of student retention, student progression and cost saving through resource optimization would be a mirage (Chen 2010).

In particular, there are worries that since neural networks are computationally intensive, hardware challenges involving processor speed and storage capacity need to be addressed. Researchers are also concerned that neural

network training take quite some time and resources. Yet another worry is theoretical concerns that ANN is a black-box concept with the user having little or no understanding of its workings coupled with the heated debate that it is a mere biologically-inspired computational concept without being a true reflection of the natural neural network that it claims to mimic. There are also the nonfunctional (quality) concerns of an EDM expert system that includes user data privacy, interoperability, security, scalability, learner mobility, among others.

An overview of the level of attention that has been given to the open issues of neural network-based EDM in higher education in the literature is presented in Table 6 and Figure 4.

As shown in both Table 4 and Figure 4, analysis of the primary studies indicates that 34.7% of researchers that researched on open issues of neural network-based EDM in higher education made efforts to identify and resolve hardware challenges such as inadequate processor capacity and storage capacity. They opined that mining valuable information from humongous amount of data with complex input parameters such as found in the education and other mission-critical application areas require a processor with higher computational power than the traditional central processing unit (CPU). Hence, research efforts now focus on developing non-von Neuman chip that is orientated toward neural parallel processing end-to-end. The research efforts have paid up as evident in the emergence of Graphics Processing Unit (GPU) and Tensor Processing Unit (TPU).

About 26.1% of researches focused on training issues of neural network. The primary studies examined posited that the complexity and duration of training a learning machine for it to be considered adequate for generalization
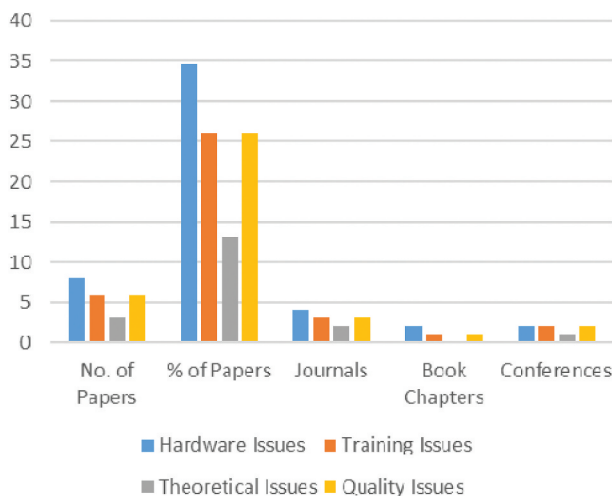


**Figure 4.** Distribution of topics on open issues of Neural Network-based EDM for HE in publication outlets.

and effectively predicting output from future input is concerning. The researchers argued that presently, training a neural network takes from hours to days. They are however hopeful that with advances in neural network hardware as highlighted above, the training period will drastically reduce.

Theoretical issues of neural networks also received moderate attention with 6 papers making 26.1% of researchers that investigated neural network-based EDM challenges. While some school of thought criticized ANN as a mere biologically-inspired computational paradigm that does not reflect the real characteristics of natural neural network, others are worried that it's black-box disposition means users only use without understanding what they are using. However, proponents of ANN responded to critics by pointing out that focus should be on the positive learning and predictive mining capacity of ANN which no other soft computing techniques like fuzzy logic, probabilistic reasoning, or evolutionary algorithm has matched.

The least discussed category of open issues are the quality concerns of predictive data mining systems like neural network-based EDM systems. The quality indicator concerns (Scheffel, Drachsler, and Specht 2015; Scheffel et al. 2014) include privacy and ethics, interoperability, and data ownership and sharing. Closely following was focus on issues of socially equitable learning experience (Wang and Jong 2016), which encompass quality indicators such as affordability, accessibility, portability and availability. In addition, researches focused on quality indicators of EDM that deal with inclusiveness since education is a social service. The authors opined that EDM systems, tools and techniques should be deployed in a manner that ensures inclusiveness.

The other quality problems that received attention in the literature though to a lesser extent, were learners' mobility, scalability, and security concerns. Efforts committed to these attributes were with a view to finding solutions. With the phenomenal growth in the use of mobile technology and its potential for promoting learner mobility and educational deepening, EDM researchers are making conscious efforts to tackle issues of resource constraints of mobile devices (Chatti et al. 2014). In any case, there is a consensus around the claims that security and scalability concerns remain thorny issues that need more attention if EDM must deliver on its purported benefits.

## Discussion

From the foregoing, we can infer that enormous research efforts have been committed to making more sense of educational data using data mining techniques (Zengin et al. 2011). As a result, advances have been made in the building of a wide range of EDM frameworks (Greller and Drachsler 2012) as platforms for data-assisted academic advising (Kraft-Terry and Kau 2016; Marsh, Pane, and Hamilton 2006; Virvou, Alepis, and Sidiropoulos 2014).

Such advisory will improve learning outcomes in higher education through prediction (Kaur, Singh, and Josan 2015), (Wanli et al. 2015) and classification (Takle and Gawai 2015), (Kaur, Singh, and Josan 2015) of learners' activities. Specific areas that have received priority attention include improving student performance (Suchithra, Vaidhehi, and Iyer 2015), identifying student's behavior (Takle and Gawai 2015), predict student performance (Shahiri, Husain, and Rashid 2015), predicting slow learners (Kaur, Singh, and Josan 2015), predicting student course selection (Ognjanovic, Gasevic, and Dawson 2016), reducing learners' course failure rate (Govindarajan, Kumar, and Boulanger 2016), and assessing course outcome (Yassine, Kadry, and Sicilia 2016). All these will culminate in boosting student retention (Gordon 2016). However, more efforts are required in the aspects of modeling student data (Palmer 2013), (Campagni et al. 2015), (Ferreira and Andrade 2014) and using more efficient learning algorithms (Wanli et al. 2015) in order to have more accurate predictions, reduce data mining time, and scale down memory space utilization when processing complex and humongous learner-related data. They are also quality problems associated with EDM systems implementation such as security, user privacy (Pardo and Siemens 2014), interoperability, scalability and learners' mobility that provide direction for future research.

Apparently, one of the predictive analytic technique fueling EDM revolution is ANN (Retalis et al. 2006),(Vahdat et al. 2015). Based on the findings of this SLR, we have acquired an overview of trends in research in the aspects of ANN modeling, learning procedures, and cost function optimization which are important to EDM researchers and practitioners alike. It is apparent that there are advances and new developments in neural network-based EDM such as network learning (Retalis et al. 2006), (Vahdat et al. 2015) and the use of ensemble methods (Wanli et al. 2015), (Budgaga et al. 2016). For instance, it was noted that neural network researchers are optimistic that a neural network-based EDM is a promising strategy that can transform the higher education sub-sector for sustainable development. Nonetheless, there is need for improved models and learning algorithms that will reduce processing time and memory utilization during data-based machine learning particularly for computationally-intensive fields like EDM. To this end, researchers are redoubling efforts to close identified gaps in all aspects of neural networks. It is worth mentioning that action research commitments overtime has led to variants of ANN models such as extreme machine learning, Boltzmann machine, deep learning, neuro-fuzzy modeling, and deep learning.

A large number of researchers are using journals and book chapters to publish their findings while others are exploring conferences and workshops as outlets for reporting their findings. Their choice of journals and book chapters as priority outlets of publication implies that the views expressed are pretty matured and well tested. Equally, the reports are complete research papers that outline reliable findings and verifiable claims. Though ANN is

already a relatively old field of research, EDM is in its formative stage. An apparent reason for the impressive interest being shown in EDM is its allied fields such as statistics, knowledge discovery in databases, data mining, data analytics, and of course, machine learning which are very developed in terms of practices, technologies, and research development. The combined effect is the offering of a platform for the quick maturity of EDM.

## Limitations

To enhance decision making and mitigate bias in the categorization process, we meticulously verified all the articles selected. This we did after the searches and the application of inclusion and exclusion criteria. Yet, we have observed that ability to have access to important papers depended on the promptness of search strings. It is possible that the search process could extract other relevant papers with a different set of search strings. Furthermore, the fact that the databases used focused more on prominent journals, books and conferences that discussed topics of neural networks and EDM implies that the opinions of lowly-rated publications that were not covered by the databases we used, were not taken into cognizance. Also, it is probable that conducting additional searches in non-English sources could have reduced bias in the work. Even after downloading some papers that were written in other languages, we could not use them owing to language barrier.

## Conclusion

EDM in higher education offers academic advising for strategic intervention in terms of adaptive learning which has pedagogical and technological implications, especially in the context of sustainable education development (Abidi et al. 2018). In spite of research efforts and advances in the fields of ANN-based EDM, more action researches are required to enhance prediction accuracy, reduce processing time, and scale down memory space utilization. This means more research work in the areas of modeling (Shum and Deakin Crick 2012), (Chatti and Schroeder 2012), (Hoel and Chen 2015a), (Hoel and Chen 2015b), educational robotics (Burbaite, Stuikys, and Damasevicius 2013) (Plauska and Damaševičius 2014) (Štuikys, Burbaite, and Damaševičius 2013), ontologies (Tankelevičiene and Damaševičius 2010) and learning algorithms (Wanli et al. 2015), 193]. For the mention, research advances have led to ensemble learning algorithms such as adaptive neuro-fuzzy inference system (ANFIS), etc. using intelligent technologies like fuzzy logic, ANN, evolutionary algorithm, and probabilistic reasoning. Robust student data modeling using such ensemble soft computing methods will ensure near-perfect data-fit. Also, learning algorithms that find optimal solution in relatively shorter time and use less memory space for training machine are expedient. With respect to

neural network modeling, learning procedures and cost function optimization, this study has unveiled the trend of research efforts in the literature. It equally revealed existing gaps that could be addressed by upcoming researchers in neural network-based EDM. Furthermore, challenges concerning neural network-based EDM have been outlined.

The critical outcomes of this study were derived from the four research questions from the view point of modeling, learning procedures, and cost function optimization. The motivation for the investigation is the desire for 1) more case studies in ANN; 2) more sophisticated implementations of neural network-based EDM expert systems; 3) the need for further studies and experiments to address hardware challenges of neural network-based EDM; 4) quest for improved models and training algorithms that map student data at close fit and predict accurately; and 5) support for ANN-based EDM system quality attributes like interoperability, privacy, security, scalability, and learners' mobility.

We are of the opinion that the picture presented in this work would be useful for researchers in ANN as well as EDM practitioners and domain experts, and administrators in higher education. They will leverage on it to make useful decisions on how best to harness the huge potentials in learning-related data using predictive data mining techniques like ANN. In the event stakeholders toe the path of neural network-based EDM, they will find the outcome of this study useful when making decisions on appropriate neural network model, learning algorithms, and cost function optimization method to adopt in any EDM task. We equally believe the study provides a veritable guide for future researchers and practitioners in the aspect of identifying research pattern, current gaps, and open issues of neural network modeling, learning algorithms, and cost function optimization.

Closing the gaps and challenges of ANN-based EDM as identified from this systematic review offers ample research direction for future researchers to make valued contributions in this relatively new domain.

## Disclosure Statement

The authors disclose no conflict of interest.

## References

Abdel-Hamid, O. et al., "Convolutional neural networks for speech recognition". IEEE/ACM Transactions on Audio, Speech, and Language Processing, Automoc, Czechoslovakia,22(10): 1533–45, 2014.

Abidi, S. M. R., M. Hussain, Y. Xu, and W. Zhang. 2018. Prediction of confusion attempting algebra homework in an intelligent tutoring system through machine learning techniques

for educational sustainable development. *Sustainability (Switzerland)* 11:1. doi:10.3390/su11010105.

Acevedo, Y. V. N., and C. E. M. Marín. 2014. System architecture based on learning analytics to educational decision makers toolkit. *Advances in Computer Science and Engineering* 13 (2):89–105.

Ali, L., M. Asadi, D. Gašević, J. Jovanovic, and M. Hatala. 2013. Factors influencing beliefs for adoption of a learning analytics tool: An empirical study. *Computers & Education* 62 (2013):130–48. doi:10.1016/j.compedu.2012.10.023.

Ali, L., M. Hatala, D. Gašević, and J. Jovanovic. 2012. A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education* 58 (1):470–89. doi:10.1016/j.compedu.2011.08.030.

Alpaydin, E. 2010. *Introduction to Machine Learning*. London: The MIT Press.

Avella, J. T., M. Kebritchi, S. G. Nunn., T. Kanai, L. Smith, and J. Meinzen-Derr. June 2016. Learning analytics methods, benefits, and challenges in higher education: a systematic literature review. *Online Learning*. 20(4):201–11. doi: 10.24059/olj.v20i4.737.

Bahadır, E. June 2016. Using neural network and logistic regression analysis to predict prospective mathematics teachers' academic success upon entering graduate education. *Educational Sciences: Theory & Practice* 16:943–64.

Barneveld, A. V., K. E. Arnold, and J. P. Campbell. 2012. Analytics in higher education: establishing a common language. *Educause Learning Initiative ELI Paper* 1.

Bassel, G. W., E. Glaab, J. Marquez, M. J. Holdsworth, and J. Bacardit. 2011. Functional network construction in arabidopsis using rule-based machine learning on large-scale data sets. *The Plant Cell* 23 (9):3101–16. doi:10.1105/tpc.111.088153.

Bengio, Y. 2012. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, ed. G. Montavon, G. B. Orr, and K. R. Müller, vol. 7700. 437–78. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-35289-8_26.

Bengio, Y., A. Courville, and P. Vincent. 2013. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8):1798–828. doi:10.1109/TPAMI.2013.50.

Bengio, Y., N. Boulanger-Lewandowski, and R. Pascanu, "Advances in optimizing recurrent networks". IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 8624–28, 2013.

Bengio, Y., R. Ducharme, P. Vincent, and C. Janvin March 2003. A neural probabilistic language model. *Journal of Machine Learning Research*. 3: 1137–55

Bernard, J., T. Chang, E. Popescu, and S. Graf, "Using artificial neural networks to identify learning styles", AIED 2015: Artificial Intelligence in Education pp 541-544, International Conference on Artificial Intelligence in Education, Madrid, Spain, 2015.

Borkar, S., and K. Rajeswari. 2014. Attributes selection for predicting students' academic performance using education data mining and artificial neural network. *International Journal of Computer Applications* 86 (10):(0975–8887). doi:10.5120/15022-3310.

Brocardo, M. L., I. Traore, I. Woungang, and M. S. Obaidat. 2017. Authorship verification using deep belief network systems. *Int J Commun Syst* 30 (12):e3259. doi:10.1002/dac.3259.

Budgaga, W., M. Malensek, S. Pallickara, N. Harvey, S. Pallickara, and S. Pallickara. 2016. Predictive analytics using statistical, learning, and ensemble methods to support real-time exploration of discrete event simulations. *Future Generation Computer Systems* 56:360–74. doi:10.1016/j.future.2015.06.013.

Burbaite, R., V. Stuikys, and R. Damasevicius (2013). Educational robots as collaborative learning objects for teaching computer science. IEEE International Conference on System Science and Engineering, London, UK, 211–16. doi:10.1109/ICSSE.2013.6614661

Campagni, R., D. Merlini, R. Sprugnoli, and M. C. Verri. 2015. Data mining models for student careers. Expert Systems with Applications. 42:5508–21.

Chatti, M. A., and U. Schroeder. 2012. Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society* 15 (3):58–76.

Chatti, M. A., V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, U., . C. Greven, A. Chakrabarti, and U. Schroeder. 2014. Learning analytics: challenges and future research directions. *eleed* 10.

Chen, C. 2010. Curricullum assessment using artificial neural networks and support vector machine modelling approaches – a case study. *IR Applications* 29. December 15, 35 - 36.

Chicco, D., P. Sadowski, and P. Baldi, "Deep autoencoder neural networks for gene ontology annotation predictions", 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14. 533–40, 2014. via ACM Digital Library, California, USA.

Cho, K., B. Van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Empiricial Methods in Natural Language Processing*, ACL Anthology Team of Volunteers Doha, 1724 - 1734. Qatar: Association for Computational Linguistics.

Choi, E., A. Schuetz, W. F. Stewart, and J. Sun. 2016. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association* 112:PMID 27521897. 1067-5027.

Ciresan, D., A. Giusti, L. M. Gambardella, and J. Schmidhuber, J. "Mitosis detection in breast cancer histology images using deep neural networks". Proceedings MICCAI, Nagoya, Japan, 2013.

Cireşan, D., U. Meier, J. Masci, and J. Schmidhuber. 2011a. Multi-column deep neural network for traffic sign classification. *Neural Networks. Selected Papers from IJCNN* 32:333–38. doi:10.1016/j.neunet.2012.02.023.

Ciresan, D., U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification", 2012a IEEE Conference on Computer Vision and Pattern Recognition: 3642–49.

Ciresan, D., U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification", 2012b IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, North America, 3642–49, 2012

Ciresan, D. C., U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification". International Joint Conference on Artificial Intelligence, Barcelona, Spain, 2011a.

Cireşan, D. C., U. Meier, L. M. Gambardella, and J. Schmidhuber. 2010a. Deep, big, simple neural nets for handwritten digit recognition. *Neural Computation*. 22 (12):3207–20. 0899-7667.

Coates, A., H. Lee, and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning". International Conference on Artificial Intelligence and Statistics (AISTATS), Forth Lauderdale, USA, 2011.

Cooper, A. 2012. A brief history of analytics. *JISC CETIS Analytics Series* 1 (9): 21.

Cooper, A. R. 2013. Learning analytics interoperability - a survey of current literature and candidate standards. *Cetis* 1 (1): 2017.

Courville, A., J. Bergstra, and Y. Bengio, "A spike and slab restricted boltzmann machine", JMLR: Workshop and Conference Proceeding, Washington, USA, 15: 233–41, 2011.

Courville, A., J. Bergstra, and Y. Bengio, "Unsupervised models of images by spike-and-slab RBMs", 28th International Conference on Machine Learning, Washingthon, USA, 10. 1–8, 2012.

D. Ciresan, A. Giusti, L.M. Gambardella, and J. Schmidhuber, Juergen (2012). Pereira, F.; Burges, C. J. C.; Bottou, L.; Weinberger, K. Q., eds. Advances in neural information processing systems 25. Curran Associates, Inc. 2843–51.

Dahl, G., D. Yu, L. Deng, and A. Acero. 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 20 (1):30–42. doi:10.1109/TASL.2011.2134090.

Damaševičius, R. (2009). Analysis of academic results for informatics course improvement using association rule mining. Paper presented at the Information Systems Development: Towards a Service Provision Society, Paphos, Cyprus, 357-363. doi:10.1007/b137171_37

Dascalu, M., C. Bodea, R. I. Mogos, A. Purnus, and B. Tesila (2018). A survey on social learning analytics: Applications, challenges and importance doi:10.1007/978-3-319-73459-0_5

Deng, L., J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, and A. Acero. "Recent advances in deep learning for speech research at Microsoft". 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 8604–08.

Deng, L. 2014. Deep learning: methods and applications. *Foundations and Trends in Signal Processing* 7 (3–4):1–199. doi:10.1561/2000000039.

Deng, L., and D. Yu. 2011. Deep convex net: A scalable architecture for speech pattern classification. *Interspeech* 2011: 2285–88.

Deng, L., D. Yu, and J. Platt, "Scalable stacking and learning for building deep architectures", 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2133–36, 2012.

Deng, L., G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview (ICASSP)", 2013.

Deng, L., G. Tur, X. He, and D. Hakkani-Tür. 2012. Use of Kernel Deep Convex Networks and End-To-End Learning for Spoken Language Understanding. In IEEE Workshop on Spoken Language Technologies, December 2012, pp. 210-215.

Deng, L., O. Abdel-Hamid, and D. Yu. 2013. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26-31 May 2013, pp. 8599-8603, IEEE.

Dietz-Uhler, B., and J. E. Hurn. 2013. "Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective". *Journal of Interactive Online Learning*. Volume 12 (Number 1):1541–4914. Spring 2013.

Drachsler, H., and W. Greller, "Privacy and analytics – It's a DELICATE issue", A Checklist to establish trusted Learning Analytics, 6th Learning Analytics and Knowledge Conference 2016, April 25-29, 2016, 89–98. Edinburgh, UK. DOI: 10.1145/2883851.2883893.

Edwards, C. 2015. Growing pains for deep learning. *Communications of the ACM* 58 (7):14–16. doi:10.1145/2771283.

Elkahky, A. M., Y. Song, and X. He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In Proceedings of the 24th International Conference on World Wide Web (WWW '15). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 278–288.

Fan, B., L. Wang, F. K. Soong, and L. Xie. 2015. Photo-Real Talking Head with Deep Bidirectional LSTM. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 19-24 April 2015, South Brisbane, QLD, Australia, pp. 4884-4888, IEEE.

Ferguson, R. 2012. Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4(5/6):304–17. January 2012. doi:10.1504/IJTEL.2012.051816.

Ferreira, S. A., and A. Andrade. 2014 August. Academic analytics: Mapping the genome of the university. *IEEE Revista Iberoamericana De Technologias Del Aprendizaje*. 9(3): 98 - 105.

Fischer, and C. Igel, "Training restricted boltzmann machines: An introduction", Pattern Recognition 47, 25–39, 2014.

Forouzanfar, M., H. R. Dajani, V. Z. Groza, M. Bolic, and S. Rajan, Comparison of feed-forward neural network training algorithms for oscillometric blood pressure estimation, 4th Int. Workshop Soft Computing Applications. Arad, Romania: IEEE, 2014.

Fournier, H., R. Kop, and H. Sitlia, "The value of learning analytics to networked learning on a personal learning environment", 1st International Conference Learning Analytics and Knowledge, February 27-March 1, 2011, Banff, Alberta, Canada.

Gao, J., X. He, S. W. Yih, and L. Deng. 2014. Learning Continuous Phrase Representations for Translation Modeling. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), June 2014, Baltimore, Maryland, USA, pp. 699–709. Association for Computational Linguistics.

Gers, F. A., and J. Schmidhuber. 2001. LSTM recurrent networks learn simple context free and context sensitive languages. *IEEE TNN* 12 (6):1333–40. doi:10.1109/72.963769.

Gopnik, A. 2017. Making AI more human: Artificial intelligence has staged a revival by starting to incorporate what we know about how children learn. *Scientific American* 316 (6):60–65. doi:10.1038/scientificamerican0617-60.

Gordon, N. A. 2016. Issues in retention and attainment in Computer Science. AdvancedHE, UK: Higher Education Academy.

Govindarajan, K., V. S. Kumar, and D. Boulanger. 2016. *Kinshuk, "Learning Analytics Solution for Reducing Learners' Course Failure Rate", In 2015 IEEE Seventh International Conference on Technology for Education (T4E),* Warangal, India, 10-12 Dec. 2015. IEEE, pp. 83-90.

Grasso, F., A. Luchetta, and S. Manetti. 2018. A multi-valued neuron based complex ELM neural network. *Neural Processing Letters* 48 (1):389–401. doi:10.1007/s11063-017-9745-9.

Graupe, D. 2013. Principles of Artificial Neural Networks.3rd Edition. Singapore: World Scientific Publishing Company Co PTE Ltd, 253–74.

Graupe, D., 2016. "Deep learning neural networks: Design and case studies". World Scientific Publishing Co Inc. 57–110.

Grefenstette, T., T. Ramalho, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, and J. Agapiou. 2016. Hybrid computing using a neural network with dynamic external memory. *Nature* 1476-4687. 538 (7626):471–76. doi:10.1038/nature20101.

Greller, W., and H. Drachsler. 2012. Translating learning into numbers: toward a generic framework for learning analytics. *Educational Technology and Society* 15 (3):42–57.

Guenaga, M., and P. Garaizar. 2016 August. From analysis to improvement: Challenges and opportunities for learning analytics. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje 11(3): 146 - 147.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning. New York: Springer-Verlag.

Hinton, G., L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, et al. 2012b. Deep neural networks for acoustic modeling in speech recognition — The shared views of four research groups. *IEEE Signal Processing Magazine*. 29(6):82–97. doi:10.1109/MSP.2012.2205597.

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. Sainath. 2012a. Deep neural networks for acoustic modeling in speech

recognition: The Shared views of four research groups. *IEEE Signal Processing Magazine* 29 (6):82–97.

Hinton, G., and R. Salakhutdinov. 2012. A better way to pretrain deep Boltzmann machines. *Advances in Neural* 3:1–9.

Hinton, G. E. 2010. A practical guide to training restricted boltzmann machines. UTML TR. *Tech. Rep* 2010–13 29 (6).

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, (2012). "Improving neural networks by preventing co-adaptation of feature detectors".

Hoel, T., and W. Chen, "Learning analytics interoperability – Looking for low-hanging fruits" In Liu, C.-C., Ogata, H., Kong, S. C., and Kashihara, A. (Eds.) (2014), 22nd International Conference on Computers in Education, 3(1): 139–158. Nara, Japan, December 2014.

Hoel, T., and W. Chen, "Privacy in learning analytics – Implications for system architecture" — In T. Watanabe and K. Seta (Eds.) (2015b). 11th International Conference on Knowledge Management, ICKM 15 in Osaka, Japan, 4 – 6 November 2015.

Hoel, T., and W. Chen. 2015a. Privacy-driven design of Learning Analytics applications – Exploring the design space of solutions for data sharing and interoperability. In *Journal of learning analytics*, 139–158.

Höst, M., and A. Oručević-Alagić. 2011. A systematic review of research on open source software in commercial software product development. *Information and Software Technology* 53 (6):616–24. doi:10.1016/j.infsof.2010.12.009.

Huang, P.-S., X. He, J. Gao, L. Deng, A. Acero, and L. Heck. "Learning deep structured semantic models for web search using clickthrough data". Microsoft Research, 2013.

Hutchinson, B., L. Deng, and D. Yu. 2012. Tensor deep stacking networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1–15:1944–57.

Hutter, M. 2012. One Decade of Universal Artificial Intelligence. *Theoretical Foundations of Artificial General Intelligence*, Atlantis Thinking Machines, vol. 4I, Atlantis Press, pp. 67–88. SBN 978-94-91216-61-9

Kalchbrenner, N., and P. Blunsom, "Recurrent continuous translation models", EMNLP'2013.

Karamouzis, S. T., and A. Vrettos. An artificial neural network for predicting student graduation outcomes. *World Congress on Engineering and Computer Science*. 2008WCECS 2008. 991-994. October 22-24, 2008. San Francisco, USA

Kaur, P., M. Singh, and G. S. Josan. 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science* 57:500–08. doi:10.1016/j.procs.2015.07.372.

Kitchenham, B., S. Charters, and L. Kuzniarz. August 2015. Guidelines for conducting systematic mapping studies in software engineering. *Journal of Information and Software Technology Archive*. 64(C):1–18. doi: 10.1016/j.infsof.2015.03.007.

Kobayashi, V., S. T. Mol, and G. Kismihók G. 0000. Labour market driven learning analytics. *Journal of Learning Analytics* 1 (3):207–10. doi:10.18608/jla.2014.13.24.

Komenda, M., M. Víta, C. Vaitsis, D. Schwarz, A. Pokorná, N. Zary, L. Dušek,2015. "Curriculum mapping with academic analytics in medical and healthcare education." *PLoS ONE* 10 (12):e0143748. doi:10.1371/journal.pone.0143748.

Kraft-Terry, S., and C. Kau. 2016. Manageable steps to implementing data-informed advising. NACADA Clearinghouse.

Krizhevsky, A., I. Sutskever, and G. Hinton. 2012. *ImageNet Classification with Deep Convolutional Neural Networks"*. Lake Tahoe, Nevada: NIPS 2012: Neural Information Processing Systems.

Kumar, G., and K. Kumar. 2012. The Use of Artificial-Intelligence-Based Ensembles for Intrusion Detection: A Review. In *Applied Computational Intelligence and Soft Computing*, 2012: 20. Article ID 850160. https://doi.org/10.1155/2012/850160

Langley, P. 2011b. The changing science of machine learning. *Machine Learning* 82 (3):275–79.

Li, X., and X. Wu. 2015. *Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015,* pp. 4520-4524, IEEE,doi:10.1109/ICASSP.2015.7178826

Lu, H., K. N. Plataniotis, and A. N. Venetsanopoulos. 0000. A survey of multilinear subspace learning for tensor data. *Pattern Recognition* 44 (7):1540–51. doi:10.1016/j.patcog.2011.01.004.

Macfadyen, L. P., and S. Dawson. 2012. Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society* 15 (3):149–63.

Marcus, G. 2017. Am I human?: Researchers need new ways to distinguish artificial intelligence from the natural kind. *Scientific American* 316 (3):58–63. doi:10.1038/scientificamerican0317-58.

Marsh, J. A., J. F. Pane, and L. S. Hamilton, "Making sense of data-driven decision making in education: Evidence from recent RAND Research", Santa Monica, CA, 2006.

Martines, H., Y. Bengio, and G. N. Yannakakis. 2013. Learning deep physiological models of affect. *IEEE Computational Intelligence* 8 (2):20–33. doi:10.1109/MCI.2013.2247823.

Mesnil, G., Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, and G. Zweig. "Using recurrent neural networks for slot filling in spoken language understanding". IEEE Transactions on Audio, Speech, and Language Processing. 23 (3): 530–39, 2015.

Mikolov, T., Karafiát, M., Burget, L., Cernocký, J. & Khudanpur, S. (2010). Recurrent neural network based language model. In T. Kobayashi, K. Hirose & S. Nakamura (eds.), INTERSPEECH ,1045-1048, ISCA.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al. 2015. Human-level control through deep reinforcement learning. *Nature*. 518(7540):529–33. doi:10.1038/nature14236.

Mohamed, A., G. Dahl, and G. Hinton. 2012. Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing* 20 (1):14–22. doi:10.1109/TASL.2011.2109382.

Mohri, M., A. Rostamizadeh, and A. Talwalkar. 2012a. *Foundations of Machine Learning*. USA, Massachusetts: MIT Press.

Mohri, M., A. Rostamizadeh, and A. Talwalkar. 2012c. *Foundations of Machine Learning*, vol. ISBN 9780262018258. USA, Massachusetts: MIT Press.

Mu, S., S. Shibata, T. Yamamoto, S. Goto, S. Nakashima, and K. Tanaka (2020). Experimental study on learning of neural network using particle swarm optimization in predictive fuzzy for pneumatic servo system doi:10.1007/978-3-030-04946-1_32

Mukherjee, S., P. Niyogi, T. Poggio, and R. Rifkin. 2006. "Learning theory: Stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization", *Advances in Computational Mathematics*. 25:161–93.

Murphy, K. P. *Machine learning: a probabilistic perspective*. Cambridge, MA, USA: MIT Press.

Nguyen, A., J. Yosinski, and J. Clune, "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images." arXiv:1412.1897, 2014.

Oancea, B., R. Dragoescu, and S. Ciucu, "Predicting students' results in higher education using neural networks", Conference Paper. April 2013, Applied Information And Communication Technologies, Jelgava, Latvia, 2013.

Ognjanovic, I., D. Gasevic, and S. Dawson. 2016. Using institutional data to predict student course selections in higher education. *Internet and Higher Education* 29:49–62. doi:10.1016/j.iheduc.2015.12.002.

Ojha, V. K., A. Abraham, and V. Snášel. 2017. Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence* 60:97–116.

Okewu, E., and O. Daramola (2017), "Design of a learning analytics system for academic advising in nigerian universities", IEEE International Conference on Computing, Networking and Informatics (IEEE ICCNI 2017), Lagos, Nigeria, October 29 – 31

Oudeyer, P.-Y. 2010. On the impact of robotics in behavioral and cognitive sciences: From insect navigation to human cognitive development. *IEEE Transactions on Autonomous Mental Development* 2 (1):2–16. doi:10.1109/TAMD.2009.2039057.

Palmer, S. 2013. Modelling engineering student academic performance using academic analytics. *International Journal of Engineering Education* 29 (1):132–38.

Papamitsiou, Z., and A. Economides. 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society* 17 (4):49–64.

Pardo, A., and G. Siemens. April 2014. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*. 45(3):438–50. doi: 10.1111/bjet.12152.

Pavlin-Bernardić, N., S. Ravić, and I. P. Matić. 2016. The application of artificial neural networks in predicting children's giftedness. *Suvremena Psihologija* 19 (1):49–59. doi:10.21465/2016-SP-191-04.

Plauska, I., and R. Damaševičius. 2014. Educational robots for internet-of-things supported collaborative learning. In *Information and software technologies. ICIST 2014. communications in computer and information science*, ed. G. Dregvaite and R. Damasevicius, vol. 465, 346–58. Cham: Springer.

Poggio, T., S. Rakhlin, A. Caponnetto, and R. Rifkin. 2012. Statistical learning theory and applications. Class: 2. https://cbmm.mit.edu/sites/default/files/documents/Class01_2019.pdf

Qianyin, X., and L. Bo. 2015. Multiple evaluation models for education based on artificial neural networks. *International Journal of Hybrid Information Technology* 8 (9):1–10. doi:10.14257/ijhit.2015.8.9.01.

Qu, S., K. Li, S. Zhang, and Y. Wang. 2018. Predicting achievement of students in smart campus. *IEEE Access* 6:60264–73. doi:10.1109/ACCESS.2018.2875742.

Rajani, S. 2011. Artificial intelligence – Man or machine. *International Journal of Information Technology and Knowledge Management*. 4 (1):173–76. Archived from the original.

Retalis, S., A. Papasalouros, Y. Psaromiligkos, S. Siscos, and T. Kargidis. 2006. Proceedings of the Fifth International Conference on Networked Learning, pp. 1-8.

Ruslan, S., and T. Joshua. 1958–71. Learning with hierarchical-deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35:2012.

Sainath, T. N., A. R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR". 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Singapore, 8614–18, 2013.

Sak, H., A. Senior, and F. Beaufays (2014a). "Long short-term memory recurrent neural network architectures for large scale acoustic modeling". in Proceedings of the 15th Annual Conference of the International Speech Communication Association: Celebrating the Diversity of Spoken Languages, INTERSPEECH 2014, pp. 338–42, September 2014.

Sak, H., A. Senior, and F. Beaufays, (2014b). "Long short-term memory recurrent neural network architectures for large scale acoustic modeling".

Scheffel, M., H. Drachsler, and M. Specht, "Developing an evaluation framework of quality indicators for learning analytics", In Fifth International Conference on Learning Analytics And Knowledge (15) (pp. 16–20). Poughkeepsie, NY, USA: ACM, 2015.

Scheffel, M., H. Drachsler, S. Stoyanov, and M. Specht. 2014. Quality indicators for learning analytics. *Educational Technology and Society* 17 (4):16–20.

Schmidhuber, J. 2015a. Deep learning in neural networks: An overview. *Neural Networks* 61:85–117.

Schmidhuber, J. 2015c. Deep learning in neural networks: an overview. *Neural Networks* 61:85–117. 2015.

Schmidhuber, J. 2015d. Deep learning in neural networks: an overview. *Neural Networks* 61:85–117.

Sclater, N., (2014), "Code of practice for learning analytics." https://www.jisc.ac.uk/sites/default/files/jd0040_code_of_practice_for_learning_analytics_190515_v1.pdf

Seddon, J. 2008. *Systems thinking in the public sector: the failure of the reform regime and a manifesto for a better way.* Axminster: Triarchy Press.

Shahiri, A. M., W. Husain, and N. A. Rashid. 2015. A review on predicting student's performance using data mining techniques. *Procedia Computer Science* 72:414–22. doi:10.1016/j.procs.2015.12.157.

Shen, Y., X. He, J. Gao, L. Deng, and G. Mesnil, (2014-11-01). "A latent semantic model with convolutional-pooling structure for information retrieval". Microsoft Research, 2014.

Shum, S. B., and R. D. Deakin Crick, "Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics", In: Proc. 2nd International Conference on Learning Analytics & Knowledge, Vancouver, BC, Canada, 29 April-2 May 2012). ACM: New York. pp.92–101, 2012.

Sidhu, G., and B. Caffo. 2014. Exploiting pitcher decision-making using reinforcement learning. The Annals of Applied Statistics, Vol. 8, No. 2 (June 2014), pp. 926-955. Institute of Mathematical Statistics.

Siemens, G., D. Gasevic, C. Haythornthwaite, S. Dawson, S. B. Shum, R. Ferguson, E. Duval, K. Verbert, and R. S. J. D. Baker, "Open learning analytics: An integrated and modularized platform", 2011.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529(7587):484–89. doi:10.1038/nature16961. 0028-0836. PMID 26819042

Sin, K., and L. Muthu. July 2015. Application of big data in education data mining and learning analytics – A literature review. *ICTACT Journal On Soft Computing: Special Issue On Soft Computing Models For Big Data* 5(4):04. doi: 10.21917/ijsc.2015.0145.

Singer, N., "Privacy concerns for classdojo and other tracking apps for school children", New York Times, 4–6, 2014.

Siri, A. 2015. Predicting students' dropout at university using artificial neural networks. *Italian Journal Of Sociology Of Education* 7 (2): 225-247.

Slade, S., and P. Prinsloo. March 2013. Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist* 57(10):1510–29. doi: 10.1177/0002764213479366.

Socher, R., "Recursive deep models for semantic compositionality over a sentiment treebank". EMNLP 2013.

Socher, R., and C. Lin, "Parsing natural scenes and natural language with recursive neural networks", 26th International Conference on Machine Learning, Bellevue, Washington, USA, 2011.

Štuikys, V., R. Burbaite, and R. Damaševičius. 2013. Teaching of computer science topics using meta-programming-based glos and lego robots. *Informatics in Education* 12 (1):125–42. doi:10.15388/infedu.2013.09.

Suchithra, R., V. Vaidhehi, and N. E. Iyer. February 2015. Survey of learning analytics based on purpose and techniques for improving student performance. *International Journal of Computer Applications* 111(1):0975–8887.

Şusnea, E. 2010. Using artificial neural networks in e-learning systems. *U.P.B. Sci. Bull* 72(4): 91-100. Series C

Sutskever, I., O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks". NIPS'2014a.

Sutskever, L., O. Vinyals, and Q. Le. 2014b. Sequence to sequence learning with neural networks. *Proc. NIPS,*3104–3112.

Sze, V., Y. Chen, T. Yang, and J. Emer, "Efficient processing of deep neural networks: A tutorial and survey". 2017.

Szegedy, A. T., and D. Erhan. 2013. Deep neural networks for object detection. *Advances in Neural Information Processing Systems*, 2553–2561.

Tahmasebi, H., and A. Hezarkhani. 2012. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers & Geosciences* 42:18–27. doi:10.1016/j.cageo.2012.02.004.

Takle, P. R., and N. Gawai. 2015. Identification of student's behavior in higher education from social media by using opinion based memetic classifier. *International Journal on Recent and Innovation Trends in Computing and Communication* 3 (3): 1074 -1078.

Tankelevičiene, L., and R. Damaševičius (2010). Towards the development of genuine intelligent ontology-based e-learning systems. IEEE International Conference on Intelligent Systems, IS 2010, London, UK, 79–84. doi:10.1109/IS.2010.5548384

Tkachenko, Y. 2015. Autonomous CRM control via CLV approximation with deep reinforcement learning in discrete and continuous action space. arXiv preprint arXiv:1504.01840.

Unterthiner, T., A. Mayr, G. Klambauer, and S. Hochreiter, "Toxicity prediction using deep learning", 2015.

Vahdat, M., A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, "Advances in learning analytics and educational data mining", ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium, 22-24 April 2015.

Vapnik, V. N., and A. Y. Chervonenkis. 1971. "On the uniform convergence of relative frequencies of events to their probabilities." *Theory of Probability and Its Applications"* 16 (2):264–80. doi:10.1137/1116025.

Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol. 2010. "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. "*The Journal of Machine Learning Research* 11:3371–408.

Virvou, M., E. Alepis, and S. Sidiropoulos, "A learning analytics tool for supporting teacher decision", von der fakultät für mathematik, informatik und naturwissenschaften der, "action research and learning analytics in higher education", 6. June 2014.

Wallach, I., M. Dzamba, and A. Heifets, "AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery", 2015

Wang, T., and M. S. Y. Jong, "Towards equitable quality education for all: Are MOOCs really a way out?", Conference 20th Global Chinese Conference on Computers in Education 2016. Hong Kong: The Hong Kong Institute of Education.

Wanli, X., G. Rui, P. Eva, and G. Sean. 2015. Participation- based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behaviour* 47:168–81. doi:10.1016/j.chb.2014.09.034.

Waxman, J. A., D. Graupe, and D. W. Carley. 2010. Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network. *American Journal of Respiratory and Critical Care Medicine* 1073-449X. 181 (7):727–33. doi:10.1164/rccm.200907-1146OC.

Widrow, B., A. Greenblatt, Y. Kim, D. Park 2013. The no-prop algorithm: A new learning algorithm for multilayer neural networks. *Neural Networks* 37:182–88. doi:10.1016/j.neunet.2012.09.020. 2013.

Yassine, S., S. Kadry, and M. Sicilia. 2016. A framework for learning analytics in moodle for assessing course outcomes. *In 2016 IEEE Global Engineering Education Conference (EDUCON)*, Abu Dhabi, United Arab Emirates, pp. 261-266, doi: 10.1109/EDUCON.2016.7474563.

Yorek, N., and I. Ugulu. September 23, 2015. "A CFBPN artificial neural network model for educational qualitative data analyses: Example of students' attitudes based on Kellerts' typologies." *Academic Journals*. 10 (18):2606–16.

Yu, D., and L. Deng. 2010. Roles of pre-training and fine-tuning in context-dependent dbn-hmms for real-world speech recognition. In *NIPS workshop on deep learning and unsupervised feature learning*, Neural Information Processing Systems.

Yu, D., and L. Deng. 2014. *Automatic speech recognition: a deep learning approach*. (*Publisher: Springer*), UK: Cambridge University Press, ISBN 978-1-4471-5779-3

Yuan, G., C. Ho, and C. Lin. 2012. "Recent advances of large-scale linear classification", proc." *IEEE* 100 (9):2584–603. doi:10.1109/JPROC.2012.2188013.

Zen, H., and H. Sak. 2015. "Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis." *ICASSP* 33 (5):4470–74.

Zen, H., and S. Hasim. 2015. "Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis." *ICASSP* 694: 4470–74.

Zengin, K., N. Esgi, E. Erginer, and M. E. Aksoy. 2011. A sample study on applying data mining research techniques in educational science: Developing a more meaning of data." *Procedia, Social and Behavioral Sciences* 15:4028–32. doi:10.1016/j.sbspro.2011.04.408.

Zhong, S., Y. Liu, and Y. Liu, (2011). "Bilinear Deep Learning for Image Classification". 19th ACM International Conference on Multimedia. MM '11. New York, NY, USA: ACM: 343–52.

Zhu, W., J. Miao, and L. Qing, (2014-07-01). "Constrained extreme learning machine: A novel highly discriminative random feedforward neural network". 2014 International Joint Conference on Neural Networks (IJCNN): 800–07, Beijing, China, 2014.

Zhu, W., J. Miao, L. Qing, and G. B. Huang, "Hierarchical extreme learning machine for unsupervised representation learning". 2015 International Joint Conference on Neural Networks (IJCNN): 1–8, Killarney, Ireland.

Zissis, D., E. K. Xidias, and D. Lekkas. 2015. A cloud based architecture capable of perceiving and predicting multiple vessel behaviour. *Applied Soft Computing* 35:652–61. doi:10.1016/j.asoc.2015.07.002.