

Received June 29, 2021, accepted July 26, 2021, date of publication August 11, 2021, date of current version August 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103931

The Application of Machine Learning to Consolidate Critical Success Factors of Lean Six Sigma

ACHINTHYA D. PERERA¹, NIHAL P. JAYAMAHA¹, NIGEL P. GRIGG¹,
MARK TUNNICLIFFE¹, AND AMARDEEP SINGH²

¹School of Food and Advanced Technology, Massey University, Palmerston North 4442, New Zealand

²School of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand

Corresponding author: Achinthya D. Perera (a.perera@massey.ac.nz)

ABSTRACT Lean six sigma (LSS) is a quality improvement phenomenon that has captured the attention of the industry. Aiming at a capability level of 3.4 defects per million opportunities (Six Sigma) and efficient (lean) processes, LSS has been shown to improve business efficiency and customer satisfaction by blending the best methods from Lean and Six Sigma (SS). Many businesses have attempted to implement LSS, but not everyone has succeeded in improving the business processes to achieve expected outcomes. Hence, understanding the cause and effect relationships of the enablers of LSS, while deriving deeper insights from the functioning of the LSS strategy will be of great value for effective execution of LSS. However, there is little research on the causal mechanisms that explain how expected outcomes are caused through LSS enablers, highlighting the need for comprehensive research on this topic. LSS literature is overwhelmed by the diverse range of Critical Success Factors (CSFs) prescribed by a plethora of conceptual papers, and very few attempts have been made to harness these CSFs to a coherent theory on LSS. We fill this gap through a novel method using artificial intelligence, more specifically Natural Language Processing (NLP), with particular emphasis on cross-domain knowledge utilization to develop a parsimonious set of constructs that explain the LSS phenomenon. This model is then reconciled against published models on SS to develop a final testable model that explains how LSS elements cause quality performance, customer satisfaction, and business performance.

INDEX TERMS Artificial intelligence, critical success factors (CSFs) of LSS, lean, lean six sigma (LSS), six sigma (SS), deep neural network, word embedding, classification model.

I. INTRODUCTION

The growth of customer demand for high-quality products and services with speedy delivery, and increased competition due to globalization have forced organizations to explore profitable solutions to gain a competitive advantage [1], [2]. Organizations across the globe have embraced various business and operational strategies to optimize their productivity and customer satisfaction [3], [4]. Lean Six Sigma (LSS) is a popular process improvement methodology comprising the Japanese-inspired Lean manufacturing (LM) [5], [6], and US-inspired Six Sigma (SS) methodologies [7]–[9]. While SS seeks to improve processes to a capability level of 3.4 defects per million opportunities, LM seeks continual improvement

toward a state of perfection, resulting in little or no unnecessary wastage in a production or service delivery system [5], [10], [11]. Both methodologies have been popular since the 1980s, while their amalgamation as ‘Lean Six Sigma’ has been discussed in SCOPUS since 2000 [12]. Figure 1 shows the growth of the usage of terms since 1989. Moreover, the increase in LSS applications in industry over the past two decades is indicative of the industry’s interest in this approach [3], [13]–[15].

A. THEORETICAL UNDERPINNINGS OF THE RESEARCH

LSS is a business performance improvement methodology that focuses on customer satisfaction through operational and service excellence. In LSS, the best elements of LM and SS are said to be optimally combined within a DMAIC project approach containing LM methods. Understanding a

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

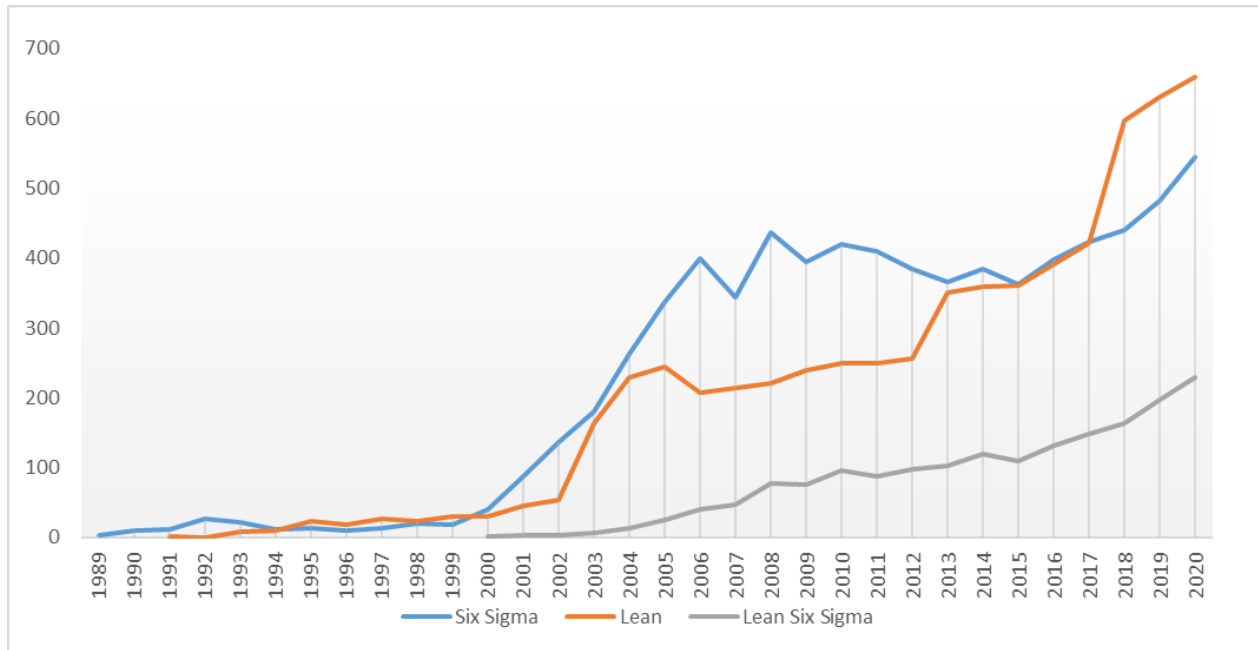


FIGURE 1. Growth of literature of terms: “Lean”, “Six Sigma” and “Lean Six Sigma”.

phenomenon such as effective implementation of SS/LSS is indeed a challenging proposition because these approaches are interwoven with the overall quality system of the organization [8], [9]. To circumvent this problem, we consider LSS as a project management approach, rather than an overall organisational management approach. By design, SS proceeds project-by-project [3], [16]–[19], allowing us to consider it as a project management methodology. Since Lean is very well incorporated in the DMAIC methodology, the LSS approach can also be viewed as a project management methodology [17]. In a project, the prime outcomes are timely completion, cost, and quality (meeting the specifications/quality goals) and as such, variables that affect project performance have been the focus in some project management literature [18], [20]. As with any other project, LSS projects also have the goal of finding a balance between the three aforesaid outcomes [17]. Without project management tools and the DMAIC process, it is difficult to ensure a balanced outcome in the LSS project, as it lacks tools to track the changes in documentation [17]. Consequently, it is important to evaluate the project aspects of LSS in order to ensure its successful deployment and sustainability.

The theoretical underpinnings of LM have been empirically tested extensively (e.g. [5], [10], [11], [21], [22]), but the same cannot be said for SS and LSS. Given the widespread use of LSS, academia and practitioners need a testable model to understand the LSS phenomenon. Some researchers (e.g. [9], [19], [23]–[25]) have attempted to explain SS as a Quality Improvement (QI) phenomenon, but the Lean element is conspicuously absent in the working definitions as well as operational definitions of these studies. While some scholars dismiss SS as a management fad [26], [27], others (e.g.: [9], [19], [23], [24], [28]) have attempted to explain SS as

a quality improvement phenomenon. Antony *et al.* [29] and Snee and Hoerl [30] pointed out that despite the evolution of SS through three decades, the theoretical basis of SS needs to be established, to which the scholars have responded to a certain extent. Key work is reviewed in Table 1.

B. BRIDGING THE LITERATURE GAPS

While the theoretical model posited by Schroeder *et al.* [9] looked promising, we found no evidence to suggest that their model has been further developed and empirically tested. One shortcoming we highlight in this model is the non-consideration of project-related contingency variables that some researchers have considered to be significant (e.g. [19], [25]). Through the thorough analysis of literature we found that no study has attempted to develop or test an explicit theory on LSS.

The second major gap we identified in the literature review is the need to integrate Lean and SS. Based on the work of Schroeder *et al.* [9] and Zu *et al.* [8], although SS does operate as a system in parallel with other improvement systems, we found that SS has not been merged with Lean in the form an explicit theoretical model. We speculate that structured problem solving [9] and Process Management [8] could be the constructs that capture both Lean and SS elements in LSS and that both constructs may be re-labelled as LSS execution.

The third significant gap we identified is that many papers have discussed the factors critical to the success (and failure) of management attempts to effectively implement Lean, SS and LSS (e.g. [28], [32], [33]), but no one has made a serious attempt to synthesize these publications to identify the broader concepts of LSS leading to LSS model development. Critical Success Factors (CSFs) of an outcome are by definition determinants, mediators and/or moderators of

TABLE 1. Literature analysis of LM and SS models.

Author	Contribution	Limitation
Linderman <i>et al.</i> [19]	Incorporated the goal theory to explain SS as a performance improvement phenomenon.	the framework does not appear to incorporate the behavioural elements of quality improvement e.g. leadership, team dynamics
Schroeder <i>et al.</i> [9]	Adopted a grounded theory approach to build their model based on fieldwork (2 Manufacturing and 2 Service); emphasise SS reporting structure ('parallel-meso structure')	They do not take into account situational factors (e.g. project complexity, project uncertainty and SS maturity) that could moderate the relationships.
Zu <i>et al.</i> [8]	Considered both the SS core and the TQM core. The hypotheses were developed from the literature and were tested empirically.	Higher model complexity (a large number of hypotheses) and the inclusion of only the manufacturing industry in the sampling frame.
Arumugam <i>et al.</i> [26]	Exploited goal theory delivering primary causal relationship between challenging goals and performance with "Knowledge" as a mediator and "Method" as a moderator.	Also considered setting challenging goals as the driver; but does not bring behavioural elements of quality improvement (e.g. leadership, culture) into play.
Sin <i>et al.</i> [25]	Used knowledge-based theory of the firm? as a basis to theorise SS project success and improving organisational performance. Used Nonaka's organizational knowledge creation theory (Nonaka, [32]).	The model does not take into account leadership as the driver of organizational performance. Did not report the operationalizations of the constructs in sufficient detail.

that outcome, and therefore research on CSFs of LSS is relevant in explaining LSS project success. Our research to date has identified 2518 separately named CSFs in the relevant literature (e.g. Abu Bakar *et al.* [34] mention 90 CSFs; Kumar *et al.* [35] mention 44 CSFs). This proliferation of factors becomes counter-productive, because it is confusing for practitioners and academics, and creates a problem in attempting to develop a parsimonious meta-model to explain the phenomenon. The same may explain why no attempt has been made to produce such a model for LSS or SS.

With so many CSFs it becomes difficult to fully 'mine' the terms in order to produce a parsimonious objective subset

of themes. Additionally, an attempt to do so might introduce the opportunity for researcher bias. Therefore this becomes a 'big data' problem that can be more effectively tackled using computing power. As a method for overcoming such problems, this paper discusses the potential of ML to confirm a parsimonious set of independent constructs through extracting the essence of CSFs leading towards a testable empirical model.

C. CRITICAL SUCCESS FACTORS OF LSS

CSFs are the elements of an organizational strategy that can influence the performance of the organization while guiding towards a positive direction [36]–[38]. Giving a boost to the search of CSF concept, Rockart [39] defined CSF as specific circumstances and variables that have a significant effect on the results and performance of an organization. In some cases, CSFs also play the role of evaluation criteria being elements that are essential for an organization or project to achieve its mission [13], [40]–[42].

Furthermore, researchers have claimed that a CSFs concept implies a direct correlation between pursuing satisfactory outcomes and performing specific activities in an organization in a specific subject area such as Continuous Improvement (CI) [38], [42], [43]. According to Alkarney and Albraithen [38] and [44], CSFs are an attempt to systematically identify the key areas that management should evaluate and prioritize when implementing LSS initiatives to achieve desired performance goals. Thus, using the CSFs concept in organizational strategic activities is vital for managers and decision-makers as it provides guidance for successful LSS implementation initiatives [43], [45].

Total quality management (TQM), Lean and SS have distinct definitions, these systems co-exist in organizations and share the same performance goals [45], [46]. In action, CI is embedded in all three systems [8], [9], [33]. Thus, in this article, we consider CSFs of Lean, Six Sigma, TQM and general CI implementation.

D. DEEP LEARNING FOR CLASSIFICATION

Artificial intelligence (AI) is general term that refers to techniques that teach machines to do things that come naturally to humans. One such AI technique is machine learning (ML), which is a set of algorithms trained on data to make decisions similar to humans. Furthermore, in ML, Deep Learning (DL) is a biological structure-inspired algorithm that mimics functions like the brain's neural structure for creating intelligent machines and systems [47], [48]. A typical supervised deep learning model consists of an input layer, which takes labelled raw data in tensor form (sometimes these could be features extracted from raw data), then works together with some hidden layers and activation functions that process input data to learn different patterns in it. Lastly, the output layer gives categorical (for classification) or real number outputs (for regression tasks). Typically a supervised learning model is trained on a large set of data until the difference between the output layer prediction and label of input difference is

minimal. In order to do this, every deep learning model has an optimization algorithm and a loss function associated with it.

DL techniques have evolved through the decades and their application has now spread not only to computer vision and autonomous vehicles but also to Natural language Processing (NLP) [47], [49]–[51]. Text classification remains a major theme in NLP, with a wide range of real-world applications in information retrieval due to its major implications [52]. In order to classify text with DL models requires input text transformed into numeric tensors. This is done by segmenting texts into tokens (words, characters etc.) and then by associating numeric vectors. There are multiple ways to vectorize or to tokenize a text. Two major ways are one-hot encoding and word embedding.

One-hot encoding uses a vocabulary index to uniquely represent a word [53], [54]. One-hot encoded vectors are binary, sparse and highly dimensional (equal to the length of vocabulary). In addition, one-hot encoding does not capture the context of a word in text, nor its semantic and syntactic similarity and relationship with other words in the text. For example, ‘similar’ and ‘same’ are related, but one-hot encoding won’t capture the semantic relationship between them. Word embedding is a popular technique that has overcome the shortcomings mentioned above. Word embedding maps a word to a vector of real numbers [55]. This mapping is not manual; in fact, the model is trained to learn weights for embedding space that project each word into vector space. There are many popular word embedding methods, and Word2Vec and GloVe are prevalent among them [47], [56]. While there is consensus that word embedding captures the semantic information of words, with regard to a full sentence it fails to capture the relationship among multiple words and phrases in the same sentence [57]. When working with textual data in NLP, for tasks which requires the meaning of a sentence, we need to consider embedding sentences or phrases. To address this, Conneau *et al.* [57] and Cer *et al.* [58] proposed sentence level embedding. This allows representing the sentence as a whole in vector form rather than combining word embedding of each word in the sentence [57], [58]. In our study, CSFs are like phrases where groups of words together make sense. Therefore, this study focuses on the using sentence embedding to vectorize CSFs for DL models.

II. METHODOLOGY

Development of the final theoretical model for a successful deployment of LSS comprises two major elements as shown in Figure 2. The first element is mining the gap in the literature by examining the existing models in both Lean and Six Sigma. The second element is to develop an ML classification model to extract the essence of CSFs in extant literature.

A. INITIAL CONCEPTUAL MODEL DEVELOPMENT

Three key basic requirements have to be full-filled in formulating a theoretical model: the establishment of the theoretical constructs, specify the relationships between the constructs based on temporal asymmetry and set the boundary within

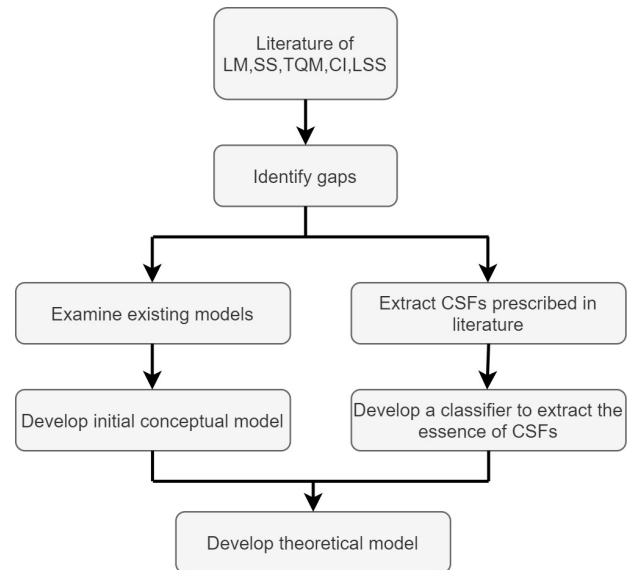


FIGURE 2. The methodology of model development.

which the theory is to be generalised [59]–[61]. Our theoretical model was built in two stages. First, we developed a baseline conceptual model (Figure 3) from the rudimentary project management axiom that maintains a project needs to be initiated and executed well to be achieve the project goals of quality, timely completion, and cost efficiency [62], [63]. This initial model was augmented to form a final theoretical model, with the aid of literature review and CSFs extraction.

Figure 3 portrays the implementation process of an LSS project. LSS project implementation comprises two major segments: the initiation phase where planning and scheduling occurs and the execution phase where conversion of process inputs (e.g. knowledge, social interactions such as teamwork, and the application tools and techniques for improvement) occurs. The success of LSS deployment is not only a product of these phases but it is a collective effort of various factors. In the LSS project initiation stage, project selection has a significant role that influences the success of LSS project deployment [64]–[66] as it aids in choosing the right project at the right time to deliver the right results. When selecting the right project, the model takes into account the four influential voices (voice of customer, voice of business, voice of process and voice of stakeholders) that were discussed in Antony and Saadat [64].

Our initial model (Figure 3) explains the positive relationship between the project initiation and LSS project execution, resulting in a successful deployment of LSS. Despite the fact Linderman *et al.* [19] has attempted to explain the initiation in terms of explicit goal setting that results in performance, there remains a gap with a question entailed “How can we achieve performance?”. Schroeder *et al.* [9] attempts to fill this gap by guiding through a structured process (DMAIC). Further, our initial model posits that successful deployment of an LSS project will result in improved quality performance and customer satisfaction, which in turn positively influences business performance [8], [67]–[69].

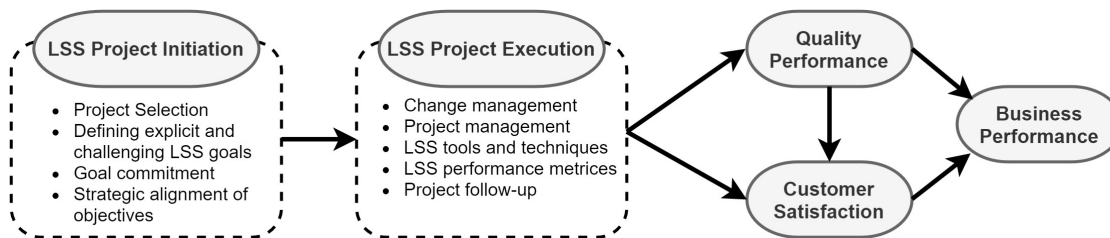


FIGURE 3. Initial conceptual model: the implementation process for the successful deployment of LSS.

B. CSFs EXTRACTION AND CLASSIFICATION

Extraction of CSFs is essential to design the classifier model using ML. We performed a literature review on CSFs for Lean, SS, TQM and CI implementation in both the manufacturing and service sectors. The literature review was carried out in the EBSCO, ELSEVIER, EMERALD, IEEE, SCOPUS and SPRINGER databases. The inclusion criteria were journal-publications from the year 2000 to present and publications that address CSFs in the areas of manufacturing and the service industry. The search included keywords such as “Critical success factors” OR “Success factors” OR “Enablers” AND “Lean Six sigma” OR “Lean Sigma” OR “Lean” OR “Six sigma” OR “LSS” OR “TQM” OR “Continuous Improvement”. Finally, 287 articles were selected for review and, and, after manual filtering, 235 articles were selected that have listed CSFs. Among the final reviewed articles, 52% represented the CSFs in the Manufacturing industry, 14% were discussing the CSFs generally (both manufacturing and service) and another 14% listed the CSFs of Small and medium enterprises (SMEs). The rest represented the CSFs of other industries such as IT, Aerospace, Construction, Education and Healthcare.

Literature analysis on CSFs revealed that prior studies related to CSFs in LSS suffer from two deficiencies. Firstly, while 3318 CSFs (from 216 articles) have been extracted from literature, only 582 (19 articles) have been classified into manageable headings (Table 3). Second, prior studies have used ad hoc methods to classify and verify their results using different methods. But, none of the studies provide a method of classifying or predicting the class for the rest of the CSFs available in the literature. As studies on CSFs are prescriptive, traditional quality criteria such as that a journal article must meet, A or A* standards, cannot be used to qualify an article on CSFs of LSS because these journals do not publish articles solely devoted to prescribing things such as CSFs. This is the reason why we used a state-of-the-art technique to sift through 3318 CSFs to be classified under manageable and meaningful themes (Leadership Engagement, LSS Culture, LSS Initiation and LSS Execution).

III. DEEP NEURAL NETWORK FOR CLASSIFICATION

Generally, Multi-class classification can be formulated as follows: $X \subset IR^D$ is a set of M instances, each of which is a D -dimensional feature vector, and C is a set of labels

TABLE 2. Attempt to classify CSFs in literature.

Citation	Industry	The method used to classify	Themes
SS [72]	M	Interpretive Structural Modelling (ISM)	11
LM [73]	H	Focused group discussion	5
SS [74]	M	Based on literature	3
SS [75]	M	Ad-hoc (validated through Partial least square)	3
LM [76]	M	Fuzzy Logic	5
SS [77]	M	Correspondence Analysis (CA)	6
LM [78]	M	Decision-Making Trial and Evaluation Laboratory (DEMATEL) analysis	3
LM [79]	M (SMEs)	Analytical Hierarchy Process (AHP)	5
CI [80]	M	Classified with ranking (Verified through Principle component analysis)	6
CI [81]	G	Authors preference	4
LSS [82]	M	Discussion with experts (Validated through fuzzy DEMATEL)	5
LSS [35]	G	Affinity Diagram process	9
LSS [83]	M	Ad-hoc (validated through Structural Equation Modelling)	7
LSS [29]	G	Fieldwork/ expert interviews	5
LSS [84]	G	Authors preference (validated through multiple-case study)	9
LSS [85]	M	Analytical Hierarchy Process (AHP)	4
LSS [86]	G	fuzzy MICMAC analysis	5
LSS [87]	G	Fuzzy Logic	4
LSS [88]	G	fuzzy AHP	5

M:Manufacturing SME: Small and Medium Enterprises
H:Healthcare
G:General

or classes [86], [87]. Each X instance is associated with a subclass of C , known as the relevant class where all other labels are irrelevant. The training model must learn a mapping function $f : IR^D \rightarrow 2^C$ that assigns a subset of class to a

given X instance in order to build a classifier. For this type of classification problem, many algorithms have been developed in the past, such as the binary significance algorithm, pairwise decomposition, and label power-set. However, in our case Deep Neural Networks (NN) scale well and function effectively by learning features from raw inputs that are typically smaller than hand-crafted features derived from raw inputs. Developing a CSF classifier is not only encompassing ML which is a subset of AI but, it is also a combination of the state-of-the-art technique NLP and subsets such as Word Embedding, Neural Network (NN) and supervised Machine Learning (SML).

To develop this Multi-Class Classifier, where the single classification label belongs to a set with more than two elements, it is essential to use frequencies of words and context data to preserve the meanings of word embedding to encode semantic significance in a word embedding [52], [88]. This study has employed TensorFlow, a Google-released open-source numerical computing framework specifically designed to ease tedious sentence embedding for the implementation of NN [58]. TensorFlow is primarily designed for creating deep NNs and it offers integrated features, such as activation, stochastic optimization techniques and convolutions for the implementation of deep learning algorithms [58], [89], [90].

A. DATA-SET DESCRIPTION

For the DL classifier we used the CSFs (from 19 articles) that were already clustered under several headings as training data. To prepare the training data, the clustered CSFs were listed with their headings. With the help of an expert panel of four, we merged several headings together and generated 4 headings (Leadership Engagement; LSS Culture; LSS Project Initiation; LSS Execution). An example of the headings merged to create the first class of Leadership Engagement is provided in Table 3.

We removed duplicate CSFs and selected only 536 unique ones with their headings in order to create the training set. The CSFs that have been examined in the Literature are generally phrases with variable length. Therefore, we implemented a universal sentence embedder which is known as the universal sentence encoder [58] from TensorFlow hub. Before encoding, we pre-processed each CSF to make all the words lowercase. The universal sentence encoder has two advantages: First, it converts single words or sentences into fixed-length vectors, which can be used for combining multiple CSFs into a single vector. This also prevents the need of padding short vectors with zeros. Since the universal encoder has been trained on a large corpus of data, it is better suited to learning with limited training data [53], [91].

During the embedding process, the training data were shuffled and the headings mapped from 1 to 4. For evaluation purposes, we split the training data set (536) into two parts: training set (80%) and cross-validation set (20%). We used stratification sampling since the training data were inconsistent, so we minimised bias in both sets of data.

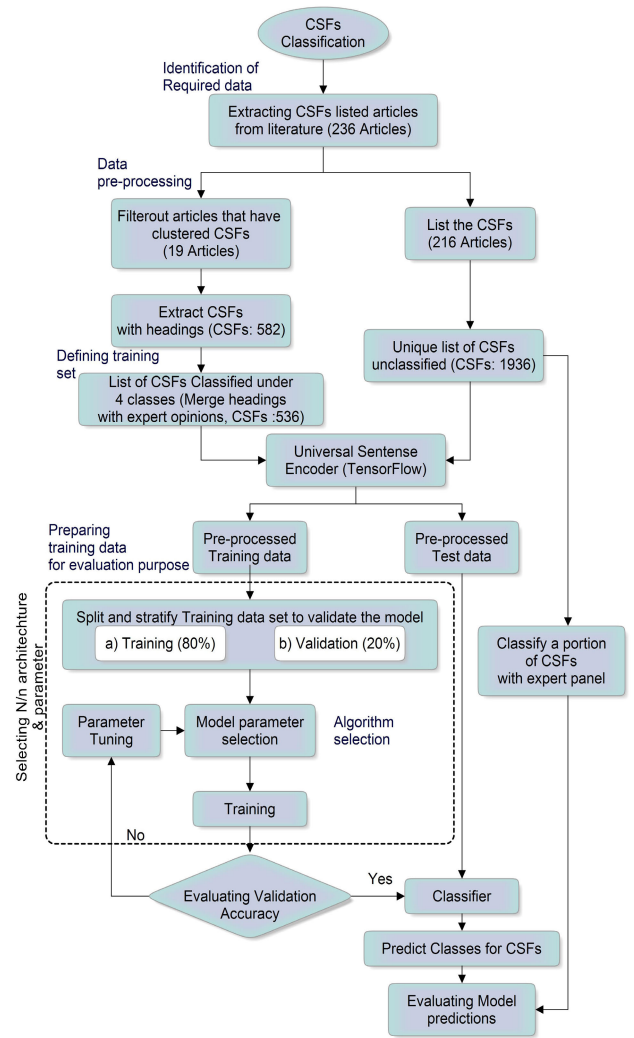


FIGURE 4. The methodology of classifier development.

B. MODEL ARCHITECTURE

The NN classification process is illustrated in Figure 5. The diagram shows a five-layer model with an input layer of 512 nodes (equal to the length of the vectored CSFs), an output layer of 4 nodes (corresponding to the number of classes), and three hidden layers of (128, 64, 64) nodes. The input layer is the first layer of the neural network, and it contains the information required for processing by subsequent layers. Our classification system also includes a sentence embedding module that has been pre-trained on a large data corpus. It converts input CSFs into a multi-dimensional (R512) vector representation for the next layer in the neural network.

For the hidden layers we employed the Rectified Linear units (ReLU) activation function $f(x) = \max(0, x)$. The dense output layer was activated before loss computation with the 'Softmax' activation parameter, which is the most general in this form of text classification task [86], [87], [92]. Softmax is a function that condenses a vector into the range of real numbers (0, 1), and all the results (probabilities) add

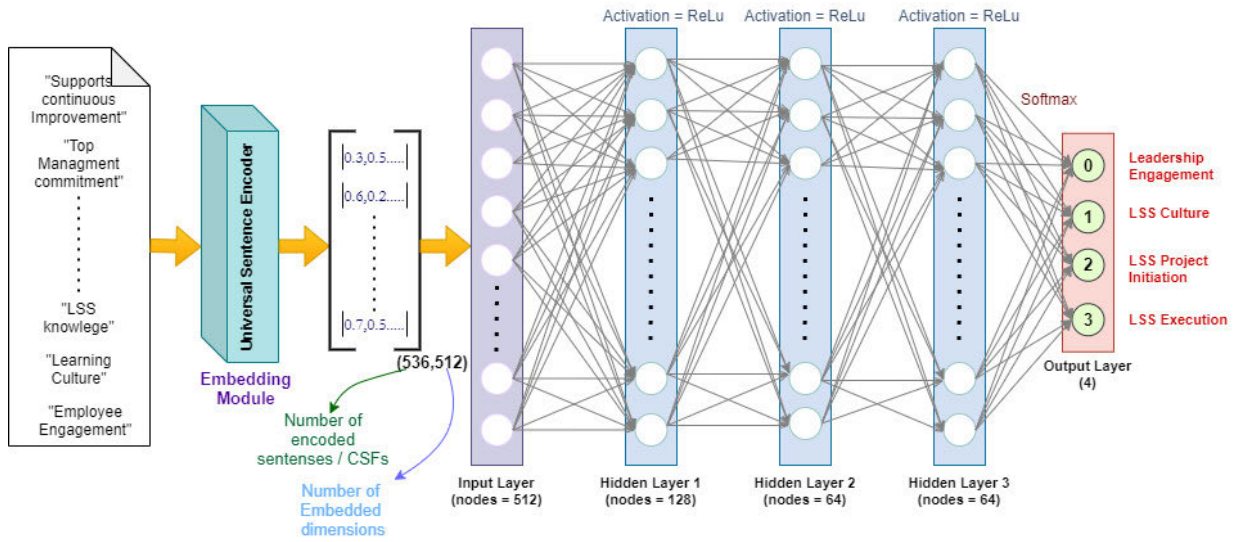


FIGURE 5. Neural network for classifier model.

up to 1 [93], [94]. Consider samples $S_i = (i = 1, 2, 3, \dots, n)$ in the training data. The Softmax function for a given classes C can be calculated as follows:

$$f(S)_i = \frac{e^{S_i}}{\sum_j^C e^{S_j}} \quad (1)$$

where S_j are the scores inferred from the total of each class in C and therefore the Softmax activation for a class S_i depends on all the scores in S . For optimizing the performance of a neural network, the cost function is critical for determining the weights of the NN [94]. After the activation function we utilised binary cross-entropy (BCE) as the loss function. BCE transforms the smoothed output from Softmax function with probabilities while penalising any deviation from the target label. The formulation for BCE can be defined as follows:

$$L = - \sum_{i=1}^2 t_i \log P_i = [t_i \log(P_i) + (1 - t_i) \log(1 - P_i)] \quad (2)$$

where, t_i is the truth value or the target label taking a value between 0 or 1 and P_i is the probability assigned from Softmax function for the i^{th} class. BCE is often considered to be the average of all data samples. Therefore, for N data points the formulation is modified as follows:

$$L = - \frac{1}{N} \cdot \sum_{i=1}^N [t_i \log(P_i) + (1 - t_i) \log(1 - P_i)] \quad (3)$$

To train a neural network, as in Figure 5, from a set X of N training instances, three steps are performed on the training stage: network initialization, parameter learning, and output activation. To avoid over-fitting, we implemented L2 regularization in hidden layers and included dropouts after each hidden layer.

C. IMPLEMENTATION DETAILS

In our model we implemented TensorFlow [89]–[91] with Keras API. Networks were trained with 100 training trials

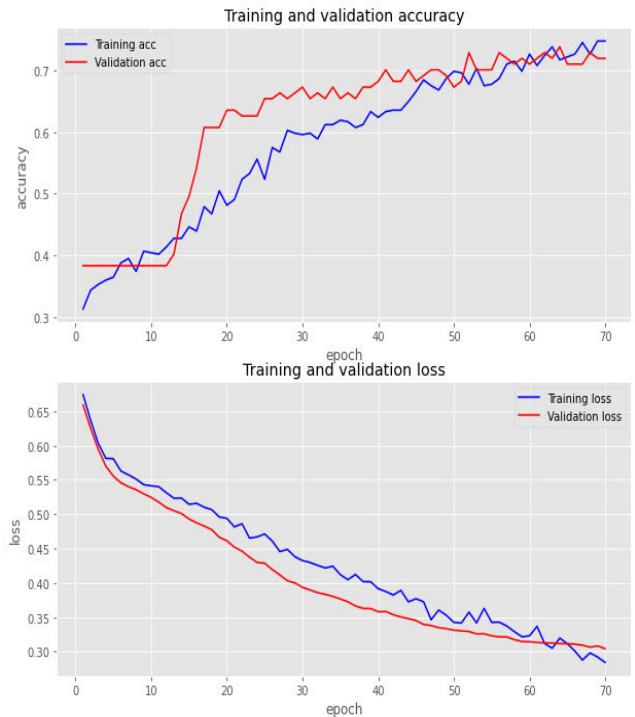


FIGURE 6. Accuracy and loss of training and cross validation.

per batch for at most 70 epochs with early stopping based on the classifier loss on the validation set. Specifically, if the validation loss for class prediction did not improve (i.e., reach a new lowest value) for 10 epochs, training was stopped and the model which resulted in the lowest validation loss was saved. Parameter updates were performed once per batch with the “rmsprop” summarized the parameter selection of our model.

D. EVALUATION METRICS

In order to evaluate the performance of the proposed method, accuracy, precision, recall and F1-scores are were chosen

TABLE 3. Merged headings to create the class: Leadership engagement.

	Heading	Critical Success Factor
Leadership Engagement	Process ownership	Top management commitment
	Human resource management	technical support; empowerment of people
	Strategic Factors	Support and commitment of Top management
	Strategic orientation	Shared vision and clear sense to lean outcomes
	Management practices	Management involvement, responsibility and commitment
	executive engagement	management must be visible and show consistent support
	Management responsibility	Management involvement, support and commitment
	Management Leadership	Commitment of the management
	Managerial	Decentralized decision-making
	Organizational	Bottom-up vs Top-down approach
	Readiness for CI deployment	leadership for CI deployment
	CI delivery and lessons learned	reward and recognition by top management
	Organization	Management involvement
	Management commitment and leadership	Strong top management involvement and commitment
	Structured improvement procedures	Managing improvement, decision making in planning process
	Focus in metric	communication on goals
	Factors relating to leadership	An organization that supports Leadership
	Organizational infrastructure	Strategic direction and alignment
	Management commitment and involvement	Support and commitment of top management
	Management commitment and involvement	Funds allocation
Strategic based LSSEs	Top-management commitment, involvement and support	
Management commitment and leadership	Leadership	
Strategy-based	clear vision and a future plan	

as the major evaluation metrics [50], [95]–[97], which are defined as follows:

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

TABLE 4. Model parameters.

Parameter		Value
Hidden layer 1	number of nodes	128
	Activation function	ReLU
	L2 regularization	0.01
	Dropout Probability	0.7
Hidden layer 2	number of nodes	64
	Activation function	ReLU
	L2 regularization	0.0001
	Dropout Probability	0.3
Hidden layer 3	number of nodes	64
	Activation function	ReLU
	L2 regularization	0.0001
	Dropout Probability	0.5
Output Layer	number of nodes	4
	Activation function	Softmax
Optimizer		rmsprop
epochs		70
Split		20%
Batch size		100

$$Recall = \frac{TP}{(TP + FN)} \tag{5}$$

$$F_1Score = \frac{(2 * precision * recall)}{(precision + recall)} \tag{6}$$

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP - FN)} \tag{7}$$

where *TP*, *FP* and *FN* are the true positives, false positives and false negatives, respectively. Precision, Recall are expedient measures of success of prediction when the classes are very imbalanced. In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. The F1-score is the average of precision and recall [47], [53].

IV. RESULTS AND DISCUSSION

A. EVALUATING CLASSIFIER MODEL

Through repeated re-sampling, cross-validation allows models to be tested using the entire training set, maximizing the total number of points used for testing and possibly to reduce the chance of over-fitting [98]. In this classification model we used 20% of the training data with designated classes as the cross validation data set. Since, there are imbalanced data for each class in the classified CSFs, we stratified the data before splitting the training set to limit the bias that can occur. Table 5, exemplifies the evaluation criteria of the classifier predictions. According to the table within the CSV set, 82.2% of the classification predictions are accurate. This level of accuracy is acceptable for our theoretical model development purpose.

The accuracy of the training model and the cross-validation model as well as the loss curves for both are depicted in Figure 6. The test data represents the 1936 CSFs extracted from literature that have not been classified. With the classifier model we generated predictions for these unclassified CSFs. For further validation and to assess the performance of

TABLE 5. Evaluation metrics for cross validation.

	precision	recall	f1-score
Leadership Engagement	0.684	0.867	0.765
LSS Culture	0.902	0.881	0.892
LSS Project Initiation	0.818	0.871	0.844
LSS Project Execution	0.786	0.579	0.667
accuracy			0.822
macro avg	0.798	0.799	0.792
weighted avg	0.827	0.822	0.820

TABLE 6. Evaluation metrics for test data.

	precision	recall	f1-score	support
Leadership Engagement	0.978	0.926	0.951	282
LSS Culture	0.992	0.946	0.968	812
LSS Project Initiation	0.943	0.981	0.962	574
LSS Project Execution	0.889	0.989	0.936	268
accuracy			0.959	1936
macro avg	0.951	0.960	0.954	1936
weighted avg	0.961	0.959	0.959	1936

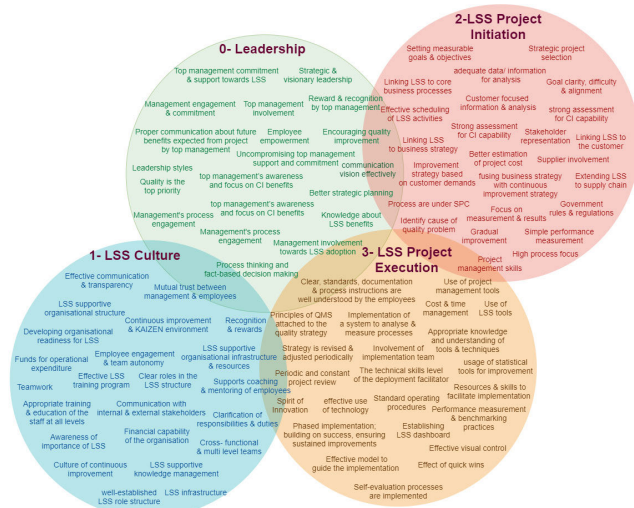


FIGURE 7. Graphical representation of training data.

this classifier we employed an expert panel of four to classify a portion of test data under the four headings. For the panel evaluation we utilized the CSFs classification in the training model as shown in Figure 7.

To evaluate the results of the expert panel assessment we used the metric root mean square error (RMSE), which provides an indication regarding the dispersion or the variability of the prediction accuracy as shown in equation 8.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (8)$$

Based on the five-point Likert scale, $RMSE = 0.78$. Since both “strongly agree” and “agree” indicate a valid prediction, we recalculated the $RMSE$ value using a three-point Likert scale. In particular, the three-point Likert scale generated

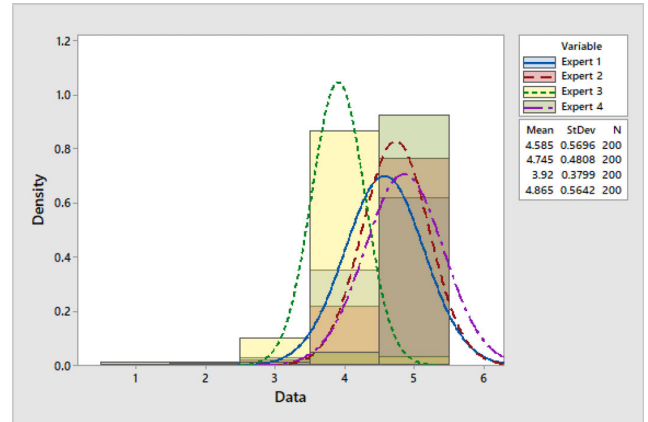


FIGURE 8. Histogram of expert panel results.

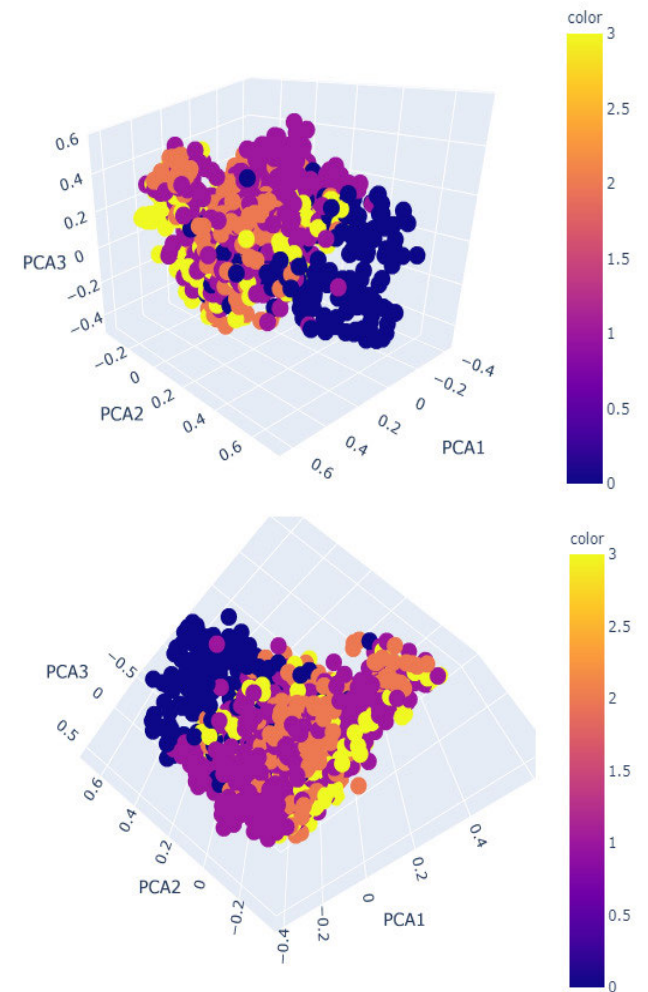


FIGURE 9. 3D scatter-plot of the features obtained from the test set.

an $RMSE$ of 0.26, indicating that our prediction model is accurate. Importantly, this measure confirms that we can rely on the manual classification as ground truth to evaluate our ML model. The histogram of how the experts responded to the prediction (Figure 8) reveals that the results are Left

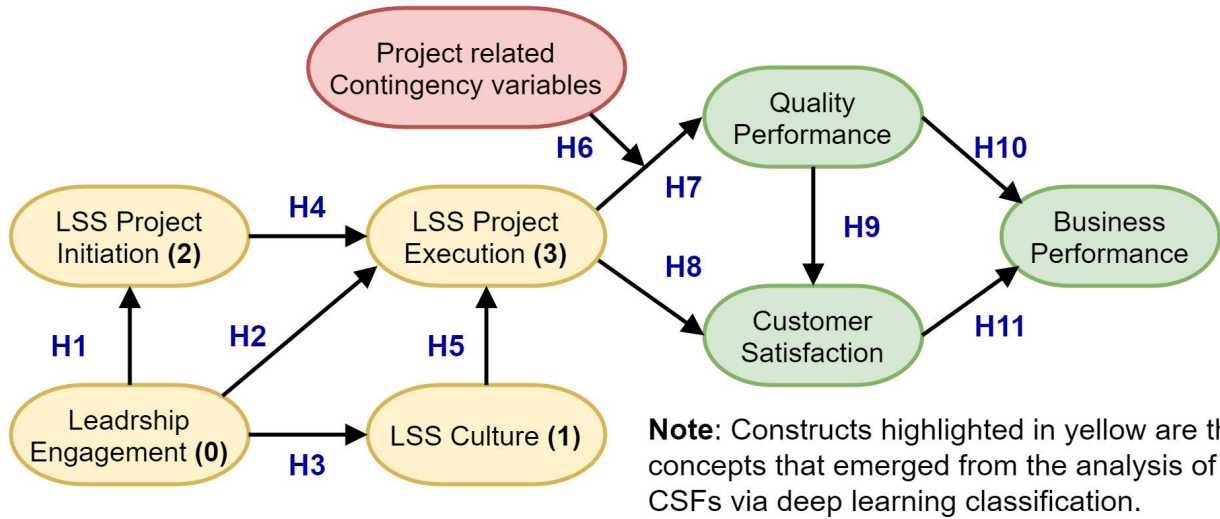


FIGURE 10. The final theoretical model explaining the LSS phenomenon.

skewed as we target the “strongly agree” response for each prediction.

A DL model was tested against the predication of the manually classified data in order to evaluate its performance according to the evaluation criteria. Table 6 depicts the evaluation metrics of test data. In order to test model predictions, the model was executed with 100 epochs and a batch size of 10. Testing accuracy was calculated to be 95.95%. Precision, recall and F1-scores are shown in the Table 6.

The CSFs in literature are sentence-based. To interpret the relationship between CSFs to each other, a higher dimensional vector space is needed because the semantic relationship between words or sentences is crucial in classification of CSFs. Principal component analysis (PCA) is a method that effectively reduces a high-dimensional vector into a lower-dimensional vector while preserving the local structure in the high-dimensional space (Figure 9). PCA can also provide greater insight into the vector space. After dimensional reduction, the word vector in the test sample appears as shown in Figure 9. On the 3D plane, rotated PCA plots indicates at least two distinct clusters (‘class 0-Leadership Engagement’ and ‘class 1-LSS Culture’). This factoid assures us that the Test data set (1936 CSFs) can be clustered under four themes.

$$ReducedDataset = FeatureVector^T \times OriginalDataset^T \quad (9)$$

B. THEORETICAL MODEL DEVELOPMENT

Figure 10 presents the constructs of the final theoretical model, and Table 7 provides an overview of the key sources that supported the constructs in the final theoretical model. These articles attempt to either build or test a parsimonious theory on SS/LSS. The analysis of the CSF literature through ML provided the researchers with greater insight into constructs that are significant but not adequately covered in SS/LSS theory building/testing studies. The relationships

TABLE 7. Key sources used to formulate theoretical constructs.

Construct	Source	Remarks
Leadership Engagement	[8], [9], [103]	Well established construct consistent with quality and also highlighted CSFs classification
LSS Project Initiation	[19]	This construct lack depth but the analysis of CSFs via ML provided the required insights
LSS Culture	[9], [19], [25], [29], [34]	Well established construct and the analysis of CSFs via ML provided additional support
LSS Execution	[8], [9]	Emerged from CSFs classification and need field work to shed more light
Project-related Contingency Variables	[19]	Need field work to identify specific variables
Quality Performance	[8], [9], [19], [25]	Well established construct consistent with quality
Customer Satisfaction	[8], [9], [19], [25]	Well established construct consistent with quality
Business Performance	[8], [9], [19], [25]	Well established construct consistent with quality

between the theoretical constructs (i.e. the hypotheses shown in Figure 10) have been derived based on existing theoretical models on SS (see Table 7). However, H9 through to H11 can be generalized across any major quality improvement initiative (for seminal work, see Deming, 1986, p.3 [99])

TABLE 8. Literature that supported the formulation of the hypotheses that constitute the model.

Hypothesis	Source
H1: Leadership Engagement has a positive effect on LSS Project Initiation.	Schroeder et al. [9]; Zu et al. [8]
H2: Leadership Engagement has a positive effect on LSS Execution.	Zu et al. [8]
H3: Leadership Engagement has a positive effect on LSS Culture.	Schroeder et al. [9]; Zu et al. [8]
H4: LSS Project Initiation has a positive effect on LSS Execution.	Schroeder et al. [9]; Zu et al. [8]
H5: LSS Culture has a positive effect on LSS Execution.	Schroeder et al. [9]; Zu et al. [8]
H6: Project-related Contingency Variables constrain the relationship between LSS Execution and Quality Performance	Lindemann et al. [26]; Needs more scoping via field work
H7: LSS Execution has a positive effect on Quality Performance.	Lindemann et al.; Schroeder et al. [9]; Sin et al. [34]; Zu et al. [8] and many others
H8: LSS Execution has a positive effect on Customer Satisfaction.	et al.; Schroeder et al. [9]; Sin et al. [34]; Zu et al. [8] and many others
H9: Quality Performance has a positive effect on Customer Satisfaction.	Lindemann et al. [26]; Schroeder et al. [9]; Sin et al. [34]; Zu et al. [8] and many others
H10: Quality Performance has a positive effect on Business Performance.	Lindemann et al. [26]; Schroeder et al. [9]; Sin et al. [34]; Zu et al. [8] and many others
H11: Customer Satisfaction has a positive effect on Business Performance.	Lindemann et al. [26]; Schroeder et al. [9]; Sin et al. [34]; Zu et al. [8] and many others

V. CONCLUSION

Despite LSS’s popularity, its theoretical underpinnings remain underdeveloped. The causal mechanisms explaining how LSS constructs are related to quality improvement project success and bottom line results of organizations remain sketchy and the Lean element of LSS is absent in almost all explanations. When explaining LSS outcomes, it is not possible to ignore CSFs since these are critical to the implementation of LSS. Moreover, the CSFs prescribed for Lean, SS, and CI are too many and poorly defined to constitute a comprehensive theory. In order to combat this problem, we presented a targeted approach for classifying CSFs that uses a deep supervised learning model to extract the essence of CSFs while supporting the development of a theoretical model. This novel approach addresses the challenges associated with the unique characteristics of the quality improvement language while extracting the essence of available literature on the CSFs related to LSS. We evaluated the effectiveness of the proposed approach to moderate

researcher bias in classifying CSFs in the literature, which is a sub-task in developing a model that explains the successful implementation of LSS. Nevertheless, further industry scoping (case research) is necessary in order to fully comprehend how to integrate Lean and SS into a theory of LSS.

Recently, NN-based approaches have shown potential and become the top classification approach for images, text, and many other databases. NNs can efficiently and automatically represent latent features as a function of the labelled training data. However, the proposed method of extracting the CSFs using a supervised deep learning-neural network is novel to the engineering management field and may open many avenues for researchers to explore. Considering the current state of the research, we intend to continue to improve the model architecture of our supervised learning model for classifying CSFs in future research work. Although the study attempts to develop and test a model while consolidating the CSFs of LSS, it has a major limitation; there is a need to conduct research in industry to “operationalize” the constructs and test statistically via a sample of LSS organisations worldwide. The proposed approach in model development is most useful to gain insights on the core concepts of practitioner driven, yet well-established methodologies (LSS is just one example). This is because the meanings of the concepts keep evolving and it is difficult reach consensus among experts on the core contests due to cognitive limitations/bounded rationality [101]. It can also be suggested that our proposed method of using Deep Learning to classify CSFs can also be applied to other fields to extract the essence of literature when it is fragmented and overloaded with terms, even in other languages.

REFERENCES

- [1] R. D. Adikorley, L. Rothenberg, and A. Guillory, “Lean six sigma applications in the textile industry: A case study,” *Int. J. Lean Six Sigma*, vol. 8, no. 2, pp. 210–224, Jun. 2017.
- [2] I. Alhuraish, C. Robledo, and A. Kobi, “A comparative exploration of lean manufacturing and six sigma in terms of their critical success factors,” *J. Cleaner Prod.*, vol. 164, pp. 325–337, Oct. 2017.
- [3] J. Antony, R. Snee, and R. Hoerl, “Lean six sigma: Yesterday, today and tomorrow,” *Int. J. Qual. Rel. Manage.*, vol. 34, no. 7, pp. 1073–1093, Aug. 2017.
- [4] R. R. Fullerton, F. A. Kennedy, and S. K. Widener, “Lean manufacturing and firm performance: The incremental contribution of lean management accounting practices,” *J. Oper. Manage.*, vol. 32, nos. 7–8, pp. 414–428, Nov. 2014.
- [5] R. Shah and P. T. Ward, “Defining and developing measures of lean production,” *J. Oper. Manage.*, vol. 25, no. 4, pp. 785–805, Jun. 2007.
- [6] H. T. S. Caldera, C. Desha, and L. Dawes, “Exploring the role of lean thinking in sustainable business practice: A systematic literature review,” *J. Cleaner Prod.*, vol. 167, pp. 1546–1565, Nov. 2017.
- [7] T. N. Goh, “A strategic assessment of six sigma,” *Qual. Rel. Eng. Int.*, vol. 18, no. 5, pp. 403–410, 2002.
- [8] X. Zu, L. D. Fredendall, and T. J. Douglas, “The evolving theory of quality management: The role of six sigma,” *J. Oper. Manage.*, vol. 26, no. 5, pp. 630–650, Sep. 2008.
- [9] R. G. Schroeder, K. Linderman, C. Liedtke, and A. S. Choo, “Six sigma: Definition and underlying theory,” *J. Oper. Manage.*, vol. 26, no. 4, pp. 536–554, Jul. 2008.
- [10] A. Reichhart and M. Holweg, “Lean distribution: Concepts, contributions, conflicts,” *Int. J. Prod. Res.*, vol. 45, no. 16, pp. 3699–3722, Aug. 2007.

- [11] R. Sundar, A. N. Balaji, and R. M. S. Kumar, "A review on lean manufacturing implementation techniques," *Procedia Eng.*, vol. 97, pp. 1875–1885, Jan. 2014.
- [12] J. H. Sheridan, "'Lean sigma' synergy," *Ind. Week*, vol. 249, no. 17, p. 81, 2000. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-8344288261&partnerID=40&md5=2f6fef46b55301e7c02afb52b9778a72d>
- [13] S. A. Albliwi, J. Antony, N. Arshed, and A. Ghadge, "Implementation of lean six sigma in Saudi Arabian organisations," *Int. J. Qual. Rel. Manage.*, vol. 34, no. 4, pp. 508–529, Apr. 2017.
- [14] L. M. Corbett, "Lean six sigma: The contribution to business excellence," *Int. J. Lean Six Sigma*, vol. 2, no. 2, pp. 118–131, May 2011.
- [15] A. Cherrafi, S. Elfezazi, A. Chiarini, A. Mokhlis, and K. Benhida, "The integration of lean manufacturing, six sigma and sustainability: A literature review and future research directions for developing a specific model," *J. Cleaner Prod.*, vol. 139, pp. 828–846, Dec. 2016.
- [16] J. de Mast, G. Diepstraten, and R. J. M. M. Does, "Quality quandaries: Design for six sigma: Method and application," *Qual. Eng.*, vol. 23, no. 2, pp. 204–211, 2011, doi: [10.1080/08982112.2011.560818](https://doi.org/10.1080/08982112.2011.560818).
- [17] B. J. Galli, "What risks does lean six sigma introduce?" *IEEE Eng. Manag. Rev.*, vol. 46, no. 1, pp. 80–90, Mar. 2018.
- [18] D. S. L. Marzagão and M. M. Carvalho, "Critical success factors for six sigma projects," *Int. J. Project Manage.*, vol. 34, no. 8, pp. 1505–1518, Nov. 2016, doi: [10.1016/j.jiproman.2016.08.005](https://doi.org/10.1016/j.jiproman.2016.08.005).
- [19] K. Linderman, R. G. Schroeder, S. Zaheer, and A. S. Choo, "Six sigma: A goal-theoretic perspective," *J. Oper. Manage.*, vol. 21, no. 2, pp. 193–203, Mar. 2003.
- [20] S. K. Gupta, A. Gunasekaran, J. Antony, S. Gupta, S. Bag, and D. Roubaud, "Systematic literature review of project failures: Current trends and scope for future research," *Comput. Ind. Eng.*, vol. 127, pp. 274–285, Jan. 2019.
- [21] R. Shah and P. T. Ward, "Lean manufacturing: Context, practice bundles, and performance," *J. Oper. Manage.*, vol. 21, no. 2, pp. 129–149, Mar. 2003.
- [22] B. R. Staats, D. J. Brunner, and D. M. Upton, "Lean principles, learning, and knowledge work: Evidence from a software services provider," *J. Oper. Manage.*, vol. 29, no. 5, pp. 376–390, Jul. 2011.
- [23] V. Arumugam, J. Antony, and K. Linderman, "A multilevel framework of six sigma: A systematic review of the literature, possible extensions, and future research," *Qual. Manage. J.*, vol. 21, no. 4, pp. 36–61, Jan. 2014.
- [24] A. B. Sin, S. Zailani, M. Iranmanesh, and T. Ramayah, "Structural equation modelling on knowledge creation in six sigma DMAIC project and its impact on organizational performance," *Int. J. Prod. Econ.*, vol. 168, pp. 105–117, Oct. 2015.
- [25] V. Arumugam, J. Antony, and K. Linderman, "The influence of challenging goals and structured method on six sigma project performance: A mediated moderation analysis," *Eur. J. Oper. Res.*, vol. 254, no. 1, pp. 202–213, Oct. 2016.
- [26] D. Näslund, "Lean, six sigma and lean sigma: Fads or real process improvement methods?" *Bus. Process Manage. J.*, vol. 14, no. 3, pp. 269–287, Jun. 2008.
- [27] M. Beer, "Why total quality management programs do not persist: The role of management quality and implications for leading a TQM transformation," *Decis. Sci.*, vol. 34, no. 4, pp. 623–642, Nov. 2003.
- [28] R. J. Hilton and A. Sohal, "A conceptual model for the successful deployment of lean six sigma," *Int. J. Qual. Rel. Manage.*, vol. 29, no. 1, pp. 54–70, Jan. 2012.
- [29] J. Antony, F. J. Antony, M. Kumar, and B. R. Cho, "Six sigma in service organisations," *Int. J. Qual. Rel. Manage.*, vol. 24, no. 3, pp. 294–311, 2007.
- [30] R. D. Snee and R. D. Hoerl, "Integrating lean and six sigma—A holistic approach," *Six Sigma Forum Mag.*, vol. 6, no. 3, pp. 15–21, 2007.
- [31] I. Nonaka, "A dynamic theory of organizational knowledge creation," *Organ. Sci.*, vol. 5, no. 1, pp. 14–37, 1994.
- [32] R. V. Sreedharan, V. Gopikumar, S. Nair, A. Chakraborty, and J. Antony, "Assessment of critical failure factors (CFFs) of lean six sigma in real life scenario," *Benchmarking, Int. J.*, vol. 25, no. 8, pp. 3320–3336, Nov. 2018.
- [33] K. Jeyaraman and L. K. Teo, "A conceptual framework for critical success factors of lean six sigma," *Int. J. Lean Six Sigma*, vol. 1, no. 3, pp. 191–215, Aug. 2010.
- [34] F. A. A. Bakar, K. Subari, and M. A. M. Daril, "Critical success factors of lean six sigma deployment: A current review," *Int. J. Lean Six Sigma*, vol. 6, no. 4, pp. 339–348, Oct. 2015.
- [35] S. Kumar, N. Kumar, and A. Haleem, "Conceptualisation of sustainable green lean six sigma: An empirical analysis," *Int. J. Bus. Excellence*, vol. 8, no. 2, pp. 210–250, 2015.
- [36] M. Maciel-Monteon, J. Limon-Romero, C. Gastelum-Acosta, D. Tlapa, Y. Baez-Lopez, and H. A. Solano-Lamphar, "Measuring critical success factors for six sigma in higher education institutions: Development and validation of a surveying instrument," *IEEE Access*, vol. 8, pp. 1813–1823, 2020.
- [37] V. F. Vallejo, J. Antony, J. A. Douglas, P. Alexander, and M. Sony, "Development of a roadmap for lean six sigma implementation and sustainability in a Scottish packing company," *TQM J.*, vol. 32, no. 6, pp. 1263–1284, Nov. 2020.
- [38] W. Alkarney and M. Albraithen, "Are critical success factors always valid for any case? A contextual perspective," *IEEE Access*, vol. 6, pp. 63496–63512, 2018.
- [39] J. F. Rockart, "Chief executives define their own data needs," *Harvard Bus. Rev.*, vol. 57, no. 2, pp. 81–93, 1979.
- [40] B. Clegg, C. Rees, and M. Titchen, "A study into the effectiveness of quality management training," *TQM J.*, vol. 22, no. 2, pp. 188–208, Mar. 2010.
- [41] A. Laureani and J. Antony, "Leadership characteristics for lean six sigma," *Total Qual. Manage. Bus. Excellence*, vol. 28, nos. 3–4, pp. 405–426, 2017.
- [42] A. R. de Jesus, J. Antony, H. A. Lepikson, and A. L. A. Peixoto, "Six sigma critical success factors in Brazilian industry," *Int. J. Qual. Rel. Manage.*, vol. 33, no. 6, pp. 702–723, Jun. 2016.
- [43] S. U. Pathiratne, A. Khatibi, and M. G. M. Johar, "CSFs for six sigma in service and manufacturing companies: An insight on literature," *Int. J. Lean Six Sigma*, vol. 9, no. 4, pp. 543–561, Oct. 2018.
- [44] J. Ram and D. Corkindale, "How 'critical' are the critical success factors (CSFs)?" *Bus. Process Manage. J.*, vol. 20, no. 1, pp. 151–174, Jan. 2014.
- [45] R. V. Sreedharan, V. M. Sunder, and R. Raju, "Critical success factors of TQM, six sigma, lean and lean six sigma," *Benchmarking, Int. J.*, vol. 25, no. 9, pp. 3479–3504, Nov. 2018.
- [46] S. Albliwi, J. Antony, S. A. H. Lim, and T. van der Wiele, "Critical failure factors of lean six sigma: A systematic literature review," *Int. J. Qual. Rel. Manage.*, vol. 31, no. 9, pp. 1012–1030, Sep. 2014.
- [47] M. Aydoğan and A. Karci, "Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 123288.
- [48] C. Liu and X. Wang, "Quality-related English text classification based on recurrent neural network," *J. Vis. Commun. Image Represent.*, vol. 71, Aug. 2020, Art. no. 102724.
- [49] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84937849144&partnerID=40&md5=6441b2a288c5fdded2adbcc8b21e092c>
- [50] Y. Zhang, H. Lin, Z. Yang, J. Wang, Y. Sun, B. Xu, and Z. Zhao, "Neural network-based approaches for biomedical relation classification: A review," *J. Biomed. Informat.*, vol. 99, Nov. 2019, Art. no. 103294. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/31557530>
- [51] P. Wang, B. Xu, J. Xu, G. Tian, C. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, pp. 806–814, Jan. 2016.
- [52] F. Gargiulo, S. Silvestri, M. Ciampi, and G. D. Pietro, "Deep neural network for hierarchical extreme multi-label text classification," *Appl. Soft Comput.*, vol. 79, pp. 125–138, Jun. 2019.
- [53] D.-H. Pham and A.-C. Le, "Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis," *Int. J. Approx. Reasoning*, vol. 103, pp. 1–10, Dec. 2018.
- [54] J. Chen, Y. Tao, and H. Lin, "Visual exploration and comparison of word embeddings," *J. Vis. Lang. Comput.*, vol. 48, pp. 178–186, Oct. 2018.
- [55] G. E. Hinton, "Learning distributed representations of concepts," in *Proc. 8th Annu. Conf. Cogn. Sci. Soc.*, vol. 1, Aug. 1986, p. 12.
- [56] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. Workshop Int. Conf. Learn. Represent. (ICLR Workshop)*, Scottsdale, AR, USA, 2013, pp. 1–12.

- [57] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," 2017, *arXiv:1705.02364*. [Online]. Available: <https://arxiv.org/abs/1705.02364>
- [58] D. Cer, Y. Yang, S.-Y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, B. Strope, and R. Kurzweil, "Universal sentence encoder for English," in *Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations*, 2018, pp. 169–174.
- [59] K. Byron and S. M. Thatcher, "Editors comments: What i know now that i wish i knew then"—Teaching theory and theory building," *Acad. Manage. Rev.*, vol. 41, no. 1, pp. 1–8, 2016.
- [60] R. Dubin, *Theory Development*. New York, NY, USA: Free Press, 1978.
- [61] D. A. Whetten, "What constitutes a theoretical contribution?" *Acad. Manage. Rev.*, vol. 14, no. 4, pp. 490–495, Oct. 1989.
- [62] R. Atkinson, "Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria," *Int. J. Project Manag.*, vol. 17, no. 6, pp. 337–342, 1999.
- [63] P. D. Gardiner and K. Stewart, "Revisiting the golden triangle of cost, time and quality: The role of NPV in project control, success and failure," *Int. J. Project Manage.*, vol. 18, no. 4, pp. 251–256, Aug. 2000.
- [64] J. Antony, "Is six sigma a management fad or fact?" *Assem. Autom.*, vol. 27, no. 1, pp. 17–19, Feb. 2007.
- [65] A. G. Psychogios, J. Antony, J. Atanasovski, and L. K. Tsironis, "Lean six sigma in a service context," *Int. J. Qual. Rel. Manage.*, vol. 29, no. 1, pp. 122–139, 2012.
- [66] R. D. Snee, "Lean six sigma—Getting better all the time," *Int. J. Lean Six Sigma*, vol. 1, no. 1, pp. 9–29, Mar. 2010.
- [67] A. Cherrafi, S. Elfezazi, K. Govindan, J. A. Garza-Reyes, K. Benhida, and A. Mokhlis, "A framework for the integration of green and lean six sigma for superior sustainability performance," *Int. J. Prod. Res.*, vol. 55, no. 15, pp. 4481–4515, Dec. 2017.
- [68] T. J. Kull and J. G. Wacker, "Quality management effectiveness in Asia: The influence of culture," *J. Oper. Manage.*, vol. 28, no. 3, pp. 223–239, May 2010.
- [69] R. Sousa and C. A. Voss, "Quality management re-visited: A reflective review and agenda for future research," *J. Oper. Manage.*, vol. 20, no. 1, pp. 91–109, Feb. 2002.
- [70] A. Talankar, P. Verma, and N. Seth, "Modelling the clusters of critical success factors of six sigma for non-formal service sectors using interpretive structural modelling," *Int. J. Six Sigma Competitive Advantage*, vol. 9, nos. 2–4, pp. 222–240, 2015, doi: [10.1504/IJSSCA.2015.074965](https://doi.org/10.1504/IJSSCA.2015.074965).
- [71] B. Noori, "The critical success factors for successful lean implementation in hospitals," *Int. J. Productiv. Qual. Manage.*, vol. 15, no. 1, pp. 108–126, 2015, doi: [10.1504/IJPM.2015.065987](https://doi.org/10.1504/IJPM.2015.065987).
- [72] Y. Z. Mehrjerdi, "Six-sigma: Methodology, tools and its future," *Assem. Autom.*, vol. 31, no. 1, pp. 79–88, Feb. 2011, doi: [10.1108/014451511111104209](https://doi.org/10.1108/014451511111104209).
- [73] N. Kumar, S. K. Jarial, and M. S. Narwal, "Assessing leanness of fast-moving consumer goods (FMCG) industry with fuzzy logic," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 2834–2840, 2019.
- [74] V. Ismyrlis and O. Moschidis, "A theoretical and statistical approach of six sigma differentiation from other quality systems," *Int. J. Lean Six Sigma*, vol. 9, no. 1, pp. 91–112, Mar. 2018, doi: [10.1108/IJLSS-07-2016-0037](https://doi.org/10.1108/IJLSS-07-2016-0037).
- [75] F. Costa, L. Lispi, A. P. Staudacher, M. Rossini, K. Kundu, and F. D. Cifone, "How to foster sustainable continuous improvement: A cause-effect relations map of lean soft practices," *Oper. Res. Perspect.*, vol. 6, 2019, Art. no. 100091, doi: [10.1016/j.orp.2018.100091](https://doi.org/10.1016/j.orp.2018.100091).
- [76] A. Belhadi, F. E. Touriki, and S. Elfezazi, "Evaluation of critical success factors (CSFs) to lean implementation in SMEs using AHP: A case study," *Int. J. Lean Six Sigma*, vol. 10, no. 3, pp. 803–829, Aug. 2019, doi: [10.1108/IJLSS-12-2016-0078](https://doi.org/10.1108/IJLSS-12-2016-0078).
- [77] V. M. Sunder and A. Prashar, "Empirical examination of critical failure factors of continuous improvement deployments: Stage-wise results and a contingency theory perspective," *Int. J. Prod. Res.*, vol. 58, no. 16, pp. 4894–4915, Aug. 2020, doi: [10.1080/00207543.2020.1727044](https://doi.org/10.1080/00207543.2020.1727044).
- [78] F. G. Aleu and E. M. Van Aken, "Systematic literature review of critical success factors for continuous improvement projects," *Int. J. Lean Six Sigma*, vol. 7, no. 3, pp. 214–232, Aug. 2016.
- [79] P. S. Parmar and T. N. Desai, "Evaluating sustainable lean six sigma enablers using fuzzy DEMATEL: A case of an Indian manufacturing organization," *J. Cleaner Prod.*, vol. 265, Aug. 2020, Art. no. 121802.
- [80] N. F. Habidin and S. M. Yusof, "Critical success factors of lean six sigma for the Malaysian automotive industry," *Int. J. Lean Six Sigma*, vol. 4, no. 1, pp. 60–82, Mar. 2013.
- [81] M. R. Mustapha, F. A. Hasan, and M. S. Muda, "Lean six sigma implementation: Multiple case studies in a developing country," *Int. J. Lean Six Sigma*, vol. 10, no. 1, pp. 523–539, Mar. 2018, doi: [10.1108/IJLSS-08-2017-0096](https://doi.org/10.1108/IJLSS-08-2017-0096).
- [82] H. Pandey, D. Garg, and S. Luthra, "Identification and ranking of enablers of green lean six sigma implementation using AHP," *Int. J. Productiv. Qual. Manage.*, vol. 23, no. 2, pp. 187–217, 2018, doi: [10.1504/IJPM.2018.089156](https://doi.org/10.1504/IJPM.2018.089156).
- [83] S. J. Raval, R. Kant, and R. Shankar, "Lean six sigma implementation: Modelling the interaction among the enablers," *Prod. Planning Control*, vol. 29, no. 12, pp. 1010–1029, Sep. 2018.
- [84] R. V. Sreedharan, R. Raju, V. M. Sunder, and J. Antony, "Assessment of lean six sigma readiness (LESIRE) for manufacturing industries using fuzzy logic," *Int. J. Qual. Rel. Manage.*, vol. 36, no. 2, pp. 137–161, Feb. 2019, doi: [10.1108/IJQRM-09-2017-0181](https://doi.org/10.1108/IJQRM-09-2017-0181).
- [85] G. Yadav and T. N. Desai, "A fuzzy AHP approach to prioritize the barriers of integrated lean six sigma," *Int. J. Qual. Rel. Manage.*, vol. 34, no. 8, pp. 1167–1185, Sep. 2017, doi: [10.1108/IJQRM-01-2016-0010](https://doi.org/10.1108/IJQRM-01-2016-0010).
- [86] J. Nam, J. Kim, E. L. Mencía, I. Gurevych, and J. Fürnkranz, "Large-scale multi-label text classification—Revisiting neural networks," in *Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer, 2014, pp. 437–452.
- [87] C. Limberg, H. Wersing, and H. Ritter, "Beyond cross-validation—Accuracy estimation for incremental and active learning models," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 3, pp. 327–346, Sep. 2020. [Online]. Available: <https://www.mdpi.com/2504-4990/2/3/18>
- [88] M. Naili, H. Chaibi, A. Ghezala, and H. H. Ben, "Comparative study of word embedding methods in topic segmentation comparative study of word embedding methods in topic segmentation," in *Proc. Int. Conf. Knowl. Based Intell. Inf. Eng. Syst.*, vol. 112, 2017, pp. 340–349.
- [89] T. Ganegedara, *Natural Language Processing With TensorFlow* (Teach Language to Machines Using Python's Deep Learning Library). 1st ed. Birmingham, U.K.: Packt Publishing Ltd., 2018.
- [90] B. Bengfort, R. Bilbro, and T. Ojeda, *Applied Text Analysis With Python*, 1st ed. Newton, MA, USA: O'Reilly Media, Inc., 2016.
- [91] S.-Y. Chen, "Use of neural network supervised learning to enhance the light environment adaptation ability and validity of green BIM," *Comput.-Aided Des. Appl.*, vol. 15, no. 6, pp. 831–840, Nov. 2018.
- [92] G. Castaneda, P. Morris, and T. M. Khoshgoftaar, "Evaluation of maxout activations in deep learning across several big data domains," *J. Big Data*, vol. 6, no. 1, pp. 1–35, Dec. 2019.
- [93] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, Aug. 2020.
- [94] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
- [95] B. Ferdinandy, L. Gerencsér, L. Corrieri, P. Perez, D. Újváry, G. Csizmadia, and Á. Miklósi, "Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures," *PLoS ONE*, vol. 15, no. 7, Jul. 2020, Art. no. e0236092, doi: [10.1371/journal.pone.0236092](https://doi.org/10.1371/journal.pone.0236092).
- [96] M. I. Lakkhal, H. Çevikalp, S. Escalera, and F. Ofli, "Recurrent neural networks for remote sensing image classification," *IET Comput. Vis.*, vol. 12, no. 7, pp. 1040–1045, Oct. 2018.
- [97] S. Bazrafkan, S. Thalavengal, and P. Corcoran, "An end to end deep neural network for iris segmentation in unconstrained scenarios," *Neural Netw.*, vol. 106, pp. 79–95, Oct. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S089360801830193X>
- [98] R. B. Rao, G. Fung, and R. Rosales, "On the dangers of cross-validation. An experimental evaluation," in *Proc. SIAM Int. Conf. Data Mining*, 2008, pp. 588–596, doi: [10.1137/1.9781611972788.54](https://doi.org/10.1137/1.9781611972788.54).
- [99] W. E. Deming, *Out of the Crisis*. Cambridge, MA, USA: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986.
- [100] A. Chakrabarty and T. K. Chuan, "An exploratory qualitative and quantitative analysis of six sigma in service organizations in Singapore," *Manage. Res. News*, vol. 32, no. 7, pp. 614–632, Jun. 2009, doi: [10.1108/01409170910965224](https://doi.org/10.1108/01409170910965224).
- [101] H. A. Simon, "A behavioral model of rational choice," *Quart. J. Econ.*, vol. 69, no. 1, pp. 99–118, 1955, doi: [10.2307/1884852](https://doi.org/10.2307/1884852).



ACHINTHYA D. PERERA received the B.Sc. degree in industrial engineering from the Asian Institute of Technology, Thailand, and the M.Sc. degree in engineering from the University of Peradeniya, Sri Lanka, in 2018. She is currently pursuing the Ph.D. degree with the School of Food and Advanced Technology, Massey University, New Zealand. Her current research interest includes lean six sigma theoretical model development while exploiting the new advancement in technology such as AI.



NIHAL P. JAYAMAHA received the B.S. degree in electrical and electronic engineering from the University of Peradeniya, Peradeniya, in 1982, and the M.E. degree in energy management and the Ph.D. degree in technology management from Massey University, New Zealand, in 2004 and 2009, respectively. From 1992 to 2001, he worked as a practicing electrical engineer in various capacities in electric utilities in Sri Lanka and the United Arab Emirates. Since 2009, he has been working as a Lecturer (Senior Lecturer since 2017) in quality management at Massey University. He has authored four book chapters and more than 100 articles. His research interests include quality improvement, lean management, and business results improvement. He received an Honorable Mention Award at the 2014 IEEE International Conference in Industrial Engineering and Engineering Management for the paper he presented on lean management.



NIGEL P. GRIGG received the B.Sc. degree (Hons.) in mathematics for business analysis from Glasgow College of Technology, Glasgow, U.K., the M.Sc. degree (Hons.) in quality management from the University of Paisley, Glasgow, and the Ph.D. degree in management science from the University of Strathclyde, Glasgow.

He has worked as a Cost Planner for IBM PC manufacturing and a Quality Engineer in super-alloy forging at Rolls Royce and Boeing supply chain. Since 1992, he has been a Lecturer, a Senior Lecturer, a Professor, and an Industry Consultant at Glasgow Caledonian University and Massey University, New Zealand. He is currently a Professor of quality systems at Massey University. His research interests include statistical quality management and improvement tools, business excellence, quality systems in agri-food production, and the role of culture as a mediator of quality practices. He has published over 270 academic and professional journal articles, book chapters, conference papers, and technical reports.

Prof. Grigg is a fellow of the Chartered Quality Institute, U.K. He is a Senior Member of the American Society for Quality, USA. He is a member, the Director, and the Former President of New Zealand Organization for Quality, New Zealand. He has given keynote and plenary addresses at quality conferences throughout Australasia and South East Asia. These outputs have gathered several best paper and highly commended paper awards and his doctoral thesis on statistical thinking in the food industry won the Institute of Quality Assurance (IQA) 2004 Annual National Award for best thesis related to quality. He is an Associate Editor of the *Total Quality Management and Business Excellence* journal (Routledge). He is serving on NZ National Reference Group to ISO Technical Committee 176 (ISO 9001).



MARK TUNNICLIFFE received the B.E. (Hons.) and Ph.D. degrees in mechanical engineering from the University of Canterbury, Christchurch, New Zealand.

He completed a Postdoctoral Research Fellowship at the Department of Materials and Metallurgical Engineering, Queen's University, ON, Canada, before working as a Mechanical Engineer in product development and project management at Tru-Test Ltd. (now Datamars Test), from 1996 to 2014, before joining Massey University, New Zealand, as a Senior Lecturer. He is currently a Senior Lecturer in product development and innovation management at Massey University. His research interests include new product development processes and its association with other elements of engineering management, engineering education, and the application of novel materials.



AMARDEEP SINGH received the master's degree in computer science and engineering from Guru Nanak Dev University, India. He is currently pursuing the Ph.D. degree with the School of Fundamental Sciences, Massey University, New Zealand. He is a Research Officer at Massey University. His fields of interest include brain-computer interface, signal processing, and machine learning.

...