WINTER
2022

JOURNAL OF MANAGEMENT & ENGINEERING INTEGRATION

AIEMS
VOL 15
NO 2

ISSN: 1939-7984

JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in the Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

Editorial Review Board

Reviewers

The Journal Editorial Team would like to thank the reviewers for their time and effort. The comments that we received were very constructive and detailed. They have been very helpful in our effort to continue to produce a top-quality journal. Your participation and timely response are very important for the success in providing a distinguished outlet for original articles. In this issue we continue to include Keywords, and the dates the publication was submitted and revised in an effort to achieve a higher standard for publication and increase the impact of the journal.

Edwin Sawan, Ph.D., P.E.
Editor-in-Chief

Junyong Ahn
Mohammed Ali
Gordon Arbogast
Ramazan Asmatulu
LuAnn Bean
Roelof deVries
Amelia Falcon
Andrzej Gapinski
Byul Hur
Abdelnasser Hussein
Xiaochun (Steven) Jiang
Robert Keyser
Krishna Krishnan
Jingyun Li
Ti Lin Liu
Adam Lynch
Holger Mauch
Roger Merriman
Dennis Ridley
Scott Swain
Dennis Tribby
Rick Wallace
Abram Walton
Gwen White
Xun Xu

Table of Contents

Predicting Asthma Patients' Total Cost Using Neural Networks and Linear Regression

Isabelina Nahmens*
Amani Ahmad
Thanos Gentimis
Laura Ikuma
Louisiana State University, USA
nahmens@lsu.edu

Abstract

In this paper, we analyze a population of asthma patients trying to predict the total cost of their treatment based on various demographic, clinical, and pharmacological data. We are comparing a neural network architecture with a simple linear regression using data from a healthcare insurance provider based in Louisiana. Our first focus was to explore the factors associated with the total cost of asthma treatment. Then we identified a sufficient threshold of data for which the neural networks outperform the linear regression models in terms of predictive accuracy. We showed that even with a simple Neural Network architecture, after approximately 6,000 randomly selected data points Neural Networks outperform Linear Regression almost always.

Keywords: Machine Learning**,** Neural Networks, Cost Estimation, Predictive Accuracy.

1. Introduction

Many data scientists believe that machine learning techniques specifically, applications of neural networks can replace almost all traditional general linear models proposed in Statistics. They argue that Neural Networks (NN) don't suffer from the characteristic problems of linear models, they are more general, they can admit different types of input and they almost always outperform traditional models in predictive accuracy (Aggarwal 2018). Although these arguments might be true generally, there is an overlooked fact that detracts from the usefulness of NNs, that they need a large number of training data to start producing accurate results. In this paper, we explore this idea and compare NNs to a traditional linear regression model with respect to accuracy as the number of data points in the training set increases.

Healthcare is one industry where NNs have a tremendous possibility of increasing the ability to analyze and predict outcomes. Healthcare data is culled from electronic health records (HER) systems and other databases and consists of massive datasets with multiple variables, of which few meet traditional criteria for performing regression analysis. Healthcare is one of the largest industries in the world, especially in the United States where we see the highest healthcare expenses compared to other developed countries. Unfortunately, the US Health System ranks last among eleven developed countries on measures of access, equity, quality, efficiency, and healthy lives (Davis et al., 2014). Therefore, efforts toward a healthcare system reform to enhance the healthcare delivery system and control costs are needed. NNs can assist by providing an alternative analysis to predict costs based on the information in those large datasets (Shickel et al., 2017).

Submitted December 1, 2021
Revised August 8, 2022

This is not the first paper where Machine Learning techniques are used in the study of asthma. Many papers have been written on the subject of identifying on classifying asthma through ML and sound inputs of the patients (Prasad et al 2011). Ours is to our knowledge the first attempt to predict the total cost of treatment for asthma patients with the use of structured EHR, asthma-related data, and Neural Networks.

Specifically, this paper will compare the regression methods and NNs ability to predict the cost of asthma, based on healthcare usage and other patient characteristics. The National Institutes of Health (NIH) defines asthma as a chronic lung disease that inflames and narrows the airways. It restricts the passage of air into the lungs and leads to episodes of wheezing, coughing, chest tightness, and shortness of breath. Asthma is one of the most common and costly diseases in the US with an estimated annual cost of $56 billion (Environmental Protection Agency, United States, 2013). Patients diagnosed with asthma are associated with increased utilization of primary health services such as hospitalization and emergency department visits as compared to those without asthma (Behr et al., 2016). Given the high cost of healthcare for this particular disease, there is an urgent need to accurately predict factors that increase healthcare costs using new approaches to analysis.

Although there is no cure for asthma, it can still be managed with proper treatment and prevention care. Proper asthma management particularly for high-risk patients can aid in improving the quality of life for patients with asthma and in controlling the cost associated with health resources utilization. High-risk patients are those patients that are at risk of not managing their condition, not adhering to their medication, and not following doctors' orders, which will result in adverse health outcomes. A study of a random sample of 1678 adults in Southeast Virginia found that those with asthma were more likely to report poor physical health and utilize more treatment services such as hospitalization and emergency department visits as compared to those without asthma (Behr, Diaz, & Akpinar-Elci, 2016). Consequently, asthma patients are characterized as frequent users of primary health services and represent a targeted population in care management programs. A successful investment in programs targeted at the right population of asthma patients may result in improved health outcomes, less demand for health services, and decreased cost.

There are a variety of clinical and non-clinical factors that can influence the level of patients' utilization of primary health services. Clinical factors include poor-quality care, insufficient care coordination between providers (physicians, nurses, and pharmacists), patients' illness severity, and any adverse outcomes. It's also worth mentioning that increased cost and utilization can be largely driven by patients' circumstances and behavior, such as lack of social support and patients' adherence to treatment, which are out of the healthcare providers' control. Therefore, previous studies explored non-clinical factors' (demographic and socio-economic) influence on readmission rates and other primary care services.

The use of predictive modeling has changed the way insurance policies are priced and managed. For example, machine learning-based techniques link environmental data to the number of visits to the emergency department for people with chronic asthma conditions (Bibi et al., 2002). Medical insurance company records such as claim data can be used to make accurate predictions about future health conditions in their members (Vaughn et. al., 2018). While predictive modeling has far less progress in improving medical outcomes, its use is imperative to predict health trends among patients- predicting which patients are likely to be hospitalized or require specialized medicine (Goss et al., 2002). By identifying high-risk patients, specialized preventative treatment can be suggested proactively to members.

The paper is structured as follows. Section 1 covers the introduction and motivation. The objective of the study can be found in section 2. In section 3 we cover the methodology including a description of the data, and the technique used to include the pre-processing phase. Section 4 contains our results, and we conclude the paper in section 5 where we also present some future work.

2. Study Objective

There were two objectives in this study: (a) Identify factors associated with the total cost for asthma treatment and build a prediction model using machine learning techniques, specifically NNs, to predict that cost based on various characteristics of asthma patients. (b) Compare the model to a linear regression model with a varying number of training data points with respect to predictive accuracy. A threshold after which the NNs start outperforming the regression model was identified.

3. Methodology

3.1. Data

The data source for this study was a healthcare insurance provider based in Louisiana. The company provides Medicaid or LaChip for qualified members through the state's Healthy Louisiana Program and links Medicaid-insured members to primary care providers, pharmacies, and case managers. The patients' data were de-identified, and a generic unique ID was created to link them across the various individual files.

Members with asthma were identified by using the ninth revision of the International Classification of Diseases codes (ICD-9Dx) with 493.XXX corresponding to asthma. We note that in the tenth revision (ICD-10Dx) the corresponding code is J45.XXX. The Drug NDC[1] numbers were used to identify asthma medications. Patients were eligible to be included in the study if they satisfied the following criteria: (1) Had a record of Medicaid insurance during the study period (January 1, 2015, to November 20, 2017) (2) Had a primary or secondary diagnosis of asthma and (3) Had at least one primary care claim (physician office visit) during the study period with asthma as the primary service diagnosis. Patients who were deceased during the study period, patients who had no primary care claim due to asthma during the study period, patients who are older than 75, and patients who had missing or invalid information were excluded from the study to comply with HIPAA regulations to reduce the probability of identifying the person. Data were extracted from a MicroStrategy database provided by the health insurance company. The claims files contained the following information: Primary physician office visits, pharmacy prescriptions, emergency department visits, hospital admissions, ICU admissions, and urgent care clinic visits.

3.2. Pre-processing

Our data pre-processing and a basic analysis was conducted in R and SAS. Patients with missing demographic information or socio-economic indicators (zip code) were removed from the dataset.

[1] National Drug Code: A unique 10-digit, 3-segment number. It is a universal product identifier for human drugs in the United States

Furthermore, patients with extreme billed amount values corresponding to a single claim were removed from the dataset, to avoid skewing the data. Finally, patients who had zero billed amount corresponding to a claim were also removed. The dataset then was edited to conform to standard notations, features, and numeric value presentations. After removing the non-asthma claims from the clean dataset, there were no hospital admissions or ICU admission claims. The final dataset thus consists of 9,977 asthma patients of all ages including infants and adults (0 to 73 years) and has 4,651 females and 5,326 males. A total of 1,609 patients are enrolled in case management, which translates to approximately 16% enrolment rate, and 81% of the patients were aged 18 and under.

3.3. Neural Networks

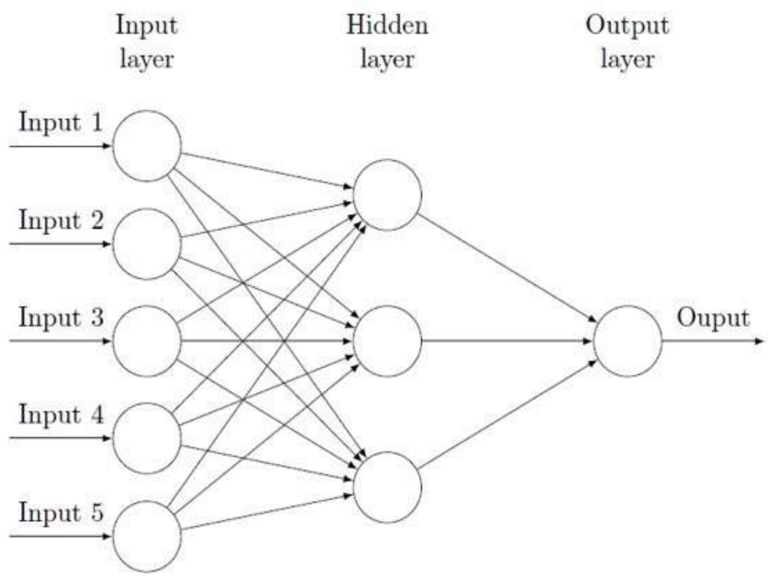


Figure 1: Neural Network with one input layer of 5 and one hidden layer of 3

In health care, NNs are used in clinical diagnosis, image analysis, and interpretation including drug development.

Neural networks (NN) can be used to classify information, predict outputs, clustering and every day new applications appear in various data-driven fields of study. NN applications include pattern recognition (Ripley 2007) image processing (Egmont-Petersen et al 2002), forecasting, classification (Zhang 2000), and others. These tools are extremely versatile and can take many types of input. For example in their paper "Classification of asthma using artificial neural networks" Badnjevic et al train a NN-based model to detect and classify asthma and present their comparisons to other methods.

Artificial Neural Networks (ANN) or just Neural Networks (Nielsen, 2015) are objectively the main tool in machine learning appropriate for handling large data sets (Figure 1). Like the human nervous system, NNs are a combination of "neurons" and "synapses" consisting of three main components: An input layer, a number of hidden layers, and an output layer. Each of these layers, which creates an n-layered NN, is connected with a set of weights and a bias factor to the next one. In addition, the combined value from each hidden layer is fed into a defined non-linear activation function, but if that function is fixed at the beginning of the analysis, only the weights and bias values will affect the output. Therefore, training a NN is a process of tuning the weights and bias values to get better accuracy through a Stochastic Gradient

Descent method. The neurons or “nodes.

4. Results

In this section, we will present some information about our data, the Neural Network architecture, and the predictive performance of the NN vs a backward-forward linear regression model. Figure 2 shows the distribution of cost which is skewed to the right as expected. This indicates that traditional statistical models may have a hard time predicting the total cost based on various features which prompted us to try non-standard methodologies through machine learning tools.

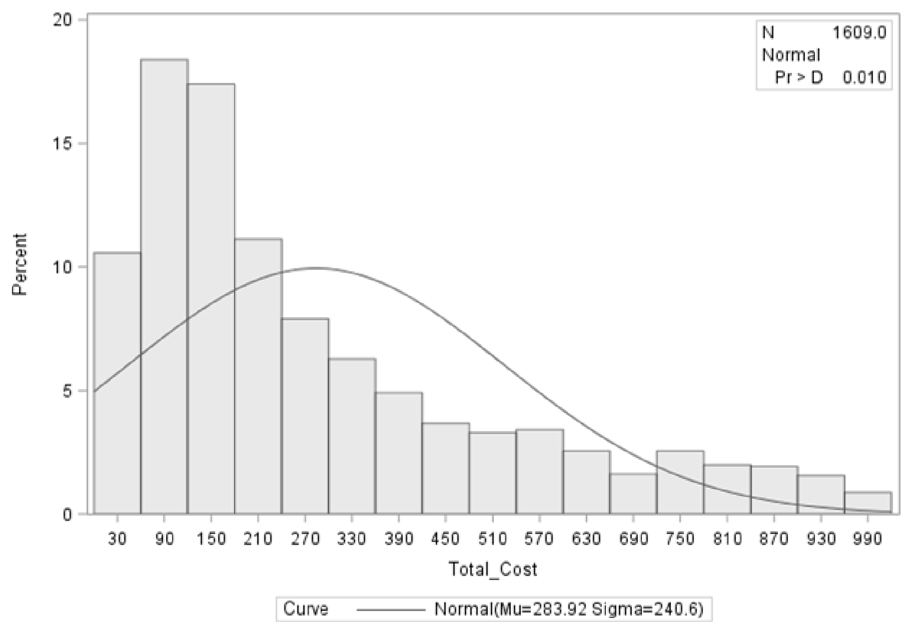


Figure 2: Distribution of Costs for the whole dataset

Generally, it is expected that a larger number of layers and nodes in a NN increases the prediction accuracy but at the cost of computational speed and memory resources. For this study and in an effort to maintain a balance between efficiency and accuracy, we chose an architecture of two hidden layers with 7 and 6 nodes each. The architecture was informed by an initial analysis of possible architectures with all possible combinations from 1 to 10 neurons in 2 layers (100 architectures) trained on randomly selected datasets of 1000 points (100 repetitions each). The 7, 6 architecture was the best on average over all iterations with just two hidden layers. We should also note here that an addition of a 3rd hidden layer with 1 to 10 neurons did not improve the accuracy of the model significantly when 1000 points were used. For a larger number of points, an increase in the accuracy might be significant enough and that will be something we explore in a subsequent paper.

Table 1: Summary of best model counts based on the number of data points

Data-points	NN	LM	Total
5,000	15	25	40
6,000	32	8	40
8,000	35	5	40

To have a robust analysis of our predictive method, we performed cross-validation with 40 random subsets of 5,000 data points. The training set had 4,500 (90%) points and then it was tested on the remaining 500 points (10%). The neural network prediction was compared to that of a multilinear regression model on the same learning and test sets. The Mean Square Error (average l2 norm) between the predicted value and the real value was used as an indicator of better prediction, for each of the subsets. We then averaged those MSEs over the number of iterations (40 repetitions). Note here that although the linear regression model always yields the same results in terms of formulas, the Neural Networks are different every time. This is due to the non-deterministic nature of Neural Networks, because of the random initialization of the weights, and the Stochastic Gradient Descent Method used to obtain the final weights. Still, the prediction results are similar for the various iterations.

For datasets of 5000 points, the average MSE between the NN prediction and true value was 1.3619. The corresponding one between the linear models' predicted value and the true value was 1.1803. Out of the 40 repetitions we tried, NN's outperformed the multilinear model only 15 times. This implies that when using 5,000 data points, for the most part, the LM is outperforming the NN in predicting the total cost.

The same process was repeated for datasets of 6,000 data points. Again, we compared the accuracy of the two models in 40 repetitions with a 90%-10% train-test ratio. This time, the NN outperformed the linear regression 32 times out of 40. So, with 6,000 data points, the NN starts outperforming the regression model.

Finally, the process was repeated for subsets of 8,000 data points, again for 40 repetitions. This time the NN outperformed the regression model 35 times. The average MSE of the neural network from the true value was 1.033 and that of the linear model was 1.317. This implies that our NN outperforms the linear model in the prediction of the total cost based on all the other input variables as the number of data points increases. Table 1 provides a summary of best model counts based on the number of data points.

The proposed threshold point based on this dataset is somewhere close to 6,000 points. It is expected that NN’s will continue to outperform linear regression as the number of data points increases. The predictive model we have created can now be used to predict the Total Cost an individual will be facing based on their characteristics, with a high enough accuracy. Figure 3 shows a visual representation of one of the networks we created.

Given a patient's demographics (age and gender) and socio-economic indicator (zip code or median income) along with a patient's history in utilizing health services in terms of the number of claims and billed amount of each claim, the NN model can attempt to predict the future total cost of the patient. This prediction can be then used to decide whether this patient should be enrolled in case management or not. Alternatively, a linear regression model can be used with similar input and output. The question of which model to use lies in the size of the available dataset. For a small number of data points, a linear model will yield more accurate predictions, but for larger data sets (around or exceeding 6,000 data points), NNs will yield more accurate predictions.

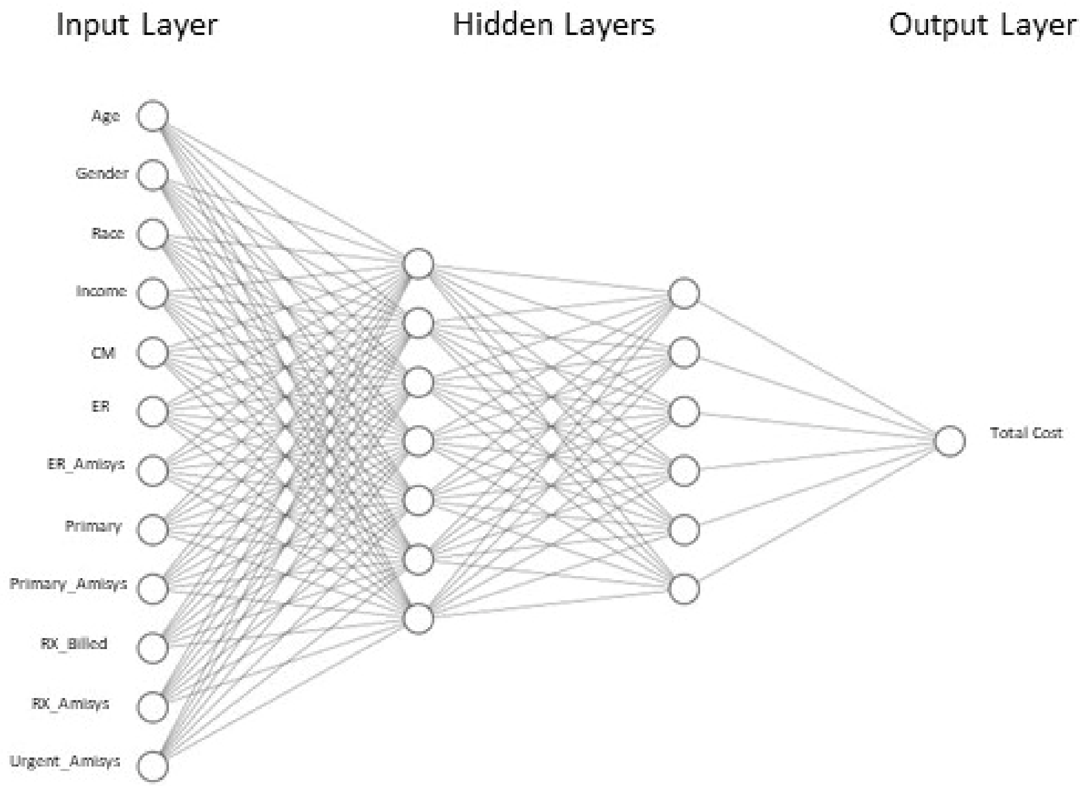


Figure 3: Visual Representation of a 7, 6 NN

5. Conclusion and Future Work

In this analysis, we primarily showed that for an Asthma total cost prediction model, the linear regression outperformed NNs in predicting total cost for the 5,000 data points run, however, the NNs started outperforming linear regression as the number of data points increased. This study thus proposes a threshold of around 6,000 points data points needed to train NNs to outperform a linear regression model. It is also expected that the larger number of data points, the better the NN will perform compared to linear regression.

This study found that asthma patients' total cost is primarily driven by the billed amount for the following claims: primary physician office visits, emergency room visits, pharmacy prescriptions, and urgent care clinic visits as was expected. Furthermore, the data analysis showed that there is a difference in total cost with respect to gender, and male patients have a higher asthma-related cost compared to female patients. Since the majority of our study population is underaged children, this finding is consistent with another study that reported that younger male patients who are Medicaid insured have higher ER visits, which translates to cost (Nath & Hsia, 2015).

With respect to factors associated with frequent utilization of primary health services, we found that patient income and enrolment in case management were the significant factors in predicting the number of emergency room visits. Similar studies that looked at the effect of gender and age on emergency room visits, reported that emergency room visits were influenced by age, where younger

patients had more visits and it decreased with age (Baibergenova et al., 2005).
For urgent care clinics visits, the patient's median income was a significant factor, patients who had higher income were more likely to utilize urgent care clinics. A previous study related socio-economic factors to another health resource utilization, and the study found that lower income is associated with a higher risk of hospitalization (Eisner et al., 2001).

There were several limitations of this study, first, patient factors such as marital status, race, and ethnicity weren't captured in the claims data. Secondly, the patient's asthma severity level was not indicated in the data which limited our ability to examine the effect of asthma severity on cost. Third, our dataset was only for Medicaid-insured patients in Louisiana, which limits the external validity of this study's findings, our findings cannot be generalized and applied to other types of insurance such as privately insured patients. And lastly, the study duration is limited, as case management service information wasn't captured in the data system before 2015. A follow-up analysis is planned, given more data in the next few years.

After this initial analysis, we proceeded to build a predictive model for the total cost, using an Artificial Neural Network through a Keras implementation in Python for subsequent years. The input variables are Age, gender, enrolment in case management program, and duration of enrolment in case management program. The information about previous years was entered including the number of emergency department claims, number of asthma medication claims, number of primary care office visits, number of urgent care clinics claims, and the billed amount for each claim for the past two years (2015-2016). This predictive model will be tested with new data being collected this year and the next one and an analysis will be presented after that.
The corresponding codes can be found in the following GitHub Repository:

7. References

Vaughn, D.A., van Deen, W. K., Kerr, W.T., Meyer, T.R., Bertozzi, A.L., Hommes, D.W., and Cohen, M.S. (2018). Using insurance claims to predict and improve hospitalizations and biologics use in members with inflammatory bowel disease. *Journal of Biomedical Informatics*, 81, 93-101.DOI:10.1016/j.jbi.2018.03.015

Prasad, B. D. C. N., Krishna Prasad, PESN, & Sagar, Y. (2011). A comparative study of machine learning algorithms as expert systems in medical diagnosis (Asthma). International Conference on Computer Science and Information Technology. Springer, Berlin, Heidelberg. DOI:10.1007/978-3-642-17857-3_56

Bibi, H., Nutman, A., Shoseyov, D., Shalom, M., Peled, R., Kivity, S. & Nutman, J. (2002). Prediction of emergency department visits for respiratory symptoms using an artificial neural network. Chest 122.5: 1627-1632. DOI:10.1378/chest.122.5.1627

Shickel, B., Tighe, P. J., Bihorac, A. & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *Journal of Biomedical and Health Informatics*, 22(5), 1589-1604. DOI:10.1109/JBHI.2017.2767063

Egmont-Petersen, M., de Ridder, D., Handels, D. (2002). Image processing with neural networks—a review, *Pattern Recognition*, 35(10), 2279-2301. DOI:10.1016/S0031-3203(01)00178-9

Ripley, B. D. (2007). *Pattern recognition and neural networks*. Cambridge university press. DOI: 10.1017/CBO9780511812651

Goss, E. P., & Vozikis, G.S. (2002). Improving health care organizational management through neural network learning. *Health Care Management Science*, 5(3), 221-227. DOI:10.1023/a:1019760901191

Nielsen, M. A. (2015). *Neural networks and deep learning*. San Francisco, CA: Determination Press.

Aggarwal, C. C. (2018). *Neural networks and deep learning* (1st ed.): Springer International Publishing.

Badnjević, A., Gurbeta, L., Cifrek, M., & Marjanovic, D. (2016). Classification of asthma using artificial neural network. The 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE. DOI:10.1109/MIPRO.2016.7522173

Moustris, K. P., Konstantinos, D., Nastos, P. T., Larissi, I. K., Anthracopoulos, M. B., Paliatsos, A. G., Priftis, K. N. (2012). Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. International journal of environmental health research 22.2: 93-104. DOI:10.1080/09603123.2011.605876

Continuous Human Pose Estimation by Machine Learning and Computer Vision

Xiaoli Huan[1*]
Hong Zhou[2]
[1]*Troy University, USA*
[2]*University of Saint Joseph, USA*
xhuan@troy.edu

Abstract

1. Introduction

Pose estimation is an important topic that has drawn significant attention from the computer vision community (Luo, Wang, Wong, & Cheng, 2020). The main task is to identify and track the pose of a person in an image or video. It can also be thought of as solving the problem of determining a set of coordinates that describe the person's pose.

Human pose estimation has a broad range of applications and is heavily used in:

a. Human-Computer Interaction: Human motion is critical to studying biomechanics and physical human-machine/robot interactions (Zhang, Chen, Yi, & Liu, 2014). Understanding the process of human locomotion can be used in clinical testing to evaluate medical disorders and affect their treatment (Simon, 2004). Pose estimation is also used in advanced driver assistance systems (ADASs) and particularly pedestrian protection systems (PPSs) to improve traffic safety (Gerónimo, López, Sappa, & Graf, 2010) and facilitate autonomous driving. Automatic arm and hand trackers for sign languages detected in videoconferencing can help disabled people to understand without requiring manual annotation (Charles, Pfister, Everingham, & Zisserman, 2014).

b. Augmented reality and gaming: Augmented reality (AR) is an interactive experience that combines real and virtual objects in a computer-generated digital environment. Pose estimation or a camera localization process solves the essential problem of registering synthetic objects seamlessly into the natural scenes (Marchand, Uchiyama, & Spindler, 2016). 6D object pose estimation is investigated in a museum augmented-reality scenario. A mobile device such as a smartphone or tablet is used to receive location and user-aware information for the exhibits (Panteleris, Michel, & Argyros, 2021). 6D pose of an object, in which 'D' refers to degrees of freedom pose, including its location and orientation, enables computers to detect and operate on the object in the correct setting (Hu, Hugonot, Fua, & Salzmann, 2019).

c. Animation in the entertainment industry: Computer-generated imagery (CGI) applies the

Submitted April 6, 2022
Revised September 4, 2022

field of computer graphics to make special effects that cannot be accomplished in reality in films and television programs. Since CGI is expensive and requires significant resources, key points can be extracted from human pose estimation in a 2D set to create 3D rendering.

Figure 1. Special effects behind the movie "Avatar" (Digital Synopsis, 2021)

d. Sports and fitness coaching: In recent years, sports data has been collected and studied to enhance an athlete's performance and reduce injury rates—machine learning and data science help use the data better and make an optimal decision. For example, pose estimation can evaluate the athletes' performance and provide detailed feedback on movements to help them achieve optimal results (Badiola-Bengoa & Mendez-Zorrilla, 2021).

2. Related Work

Pose estimation by deep learning-based methods has gained lots of interest due to its effectiveness and flexibility. Deep convolutional neural networks (CNN) can extract complex features from input data with more accurate results provided with sufficient training data. Two approaches are studied:

a) Top-down: A person is detected first, then joints and parts are estimated and calculated.

b) Bottom-up: Joints and parts are detected first, then grouped to the associated person.

The most popular deep learning-based pose estimation methods are:

1) OpenPose (Cao, Hidalgo, Simon, Wei, & Sheikh, 2021): A bottom-up approach detects parts and joints and then assigns them to persons to whom they belong. CNN is utilized as the central architecture. A VGG-19 convolutional network extracts patterns and features from the given input images. The results from the VGG-19 are assigned to the two following branches of convolutional networks. Eighteen confidence maps representing a particular part of the human pose skeleton are predicted in the first branch. The second branch forecasts a Part Affinity Fields (PAFs), which create a degree of association between parts. Weaker links in the biparties graph are adjusted, and line integrals are calculated by PAFs values to find the matching connections between parts. The remaining step is to detect connections to form final skeletons.

2) DeepCut (Pishchulin, et al., 2016): A bottom-up approach simultaneously solves the detection and pose estimation tasks. A subset of body parts such as arm, leg, and torso is selected from the possible body part candidates partitioned by CNN-based part detectors. Then, an integer linear program implicitly gathers the parts to form representations of body

parts respecting geometric and appearance constraints.

3) AlphaPose (Fang, Xie, Tai, & Lu, 2017): A top-down approach involves regional multi-person pose estimation (RMPE) two-step frameworks. The method consists of three novel components: Symmetric Spatial Transformer Network (SSTN) with parallel Single Person Pose Estimator (SPPE) to extract a high-quality single person region from an inaccurate bounding box and estimate the human pose skeleton of that person, parametric pose Non-Maximum-Suppression (NMS) takes care the problems of redundant pose deductions, and pose-guided proposals generator (PGPG) to expand training models that improve to train the SPPE and SSTN networks.
4) Mask RCNN (He, Gkioxari, Dollár, & Girshick, 2018): Mask R-CNN is based on Faster Regional based Convolutional Neural Network to implement semantic and instance segmentation. The model concurrently detects objects in the image and generates a high-quality mask that semantically segments the object. Mask R-CNN is straightforward and efficient and simply needs a small overhead to Faster R-CNN with pixel-to-pixel alignment. It can be adapted for other tasks such as human pose estimation.

3. Human Pose Estimation Implementation

Our human pose estimation utilizes MediaPipe Pose (MediaPipe, 2021), an open-source machine learning solution for accurate body pose tracking. The model has 33 landmarks and a background segmentation mask that covers the whole body from RGB video frames using the MediaPipe BlazePose (Bazarevsky, et al., 2020) research that equips with the machine learning (ML) Kit Pose Detection API (Pose Detection, 2020). In addition, the framework has built-in machine learning enhanced processing even on ordinary CPU hardware and cross platforms such as Android, iOS, desktop/cloud, web, and IoT.

The COCO topology (COCO 2020 Keypoint Detection Task, 2021), the current standard evaluating human pose technologies, consists of 17 landmarks across the face and body. However, the ankle and wrist points are localized, but the scale and orientation information is missing for hands and feet, which is crucial for gesture and fitness applications. The MediaPipe pose model includes more hands, face, or feet key points for real-time tracking.

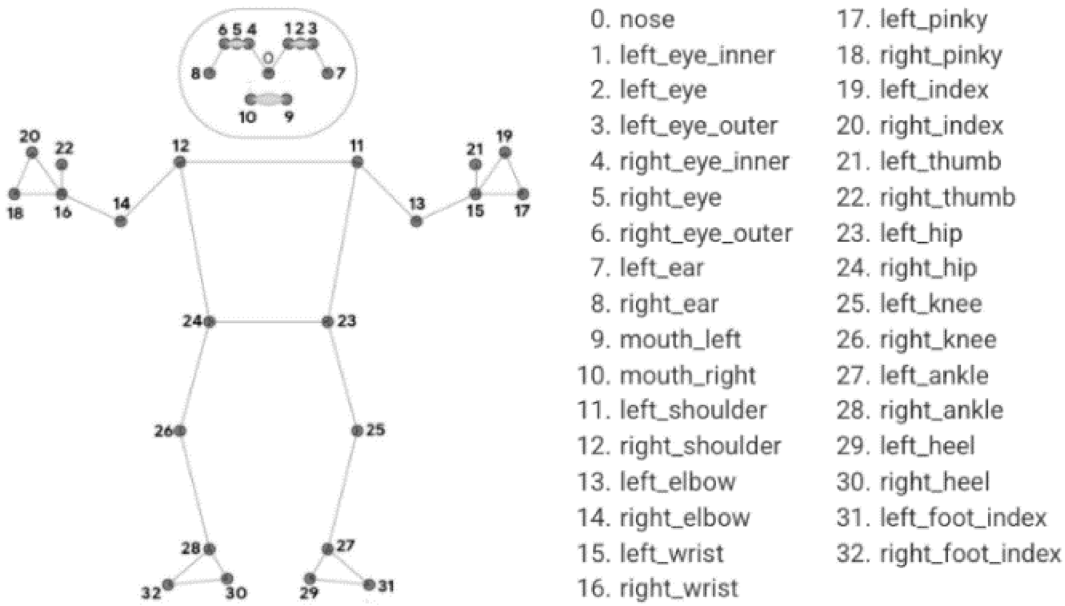


Figure 2. MediaPipe keypoints representation of a human body (MediaPipe, 2021)

A Two-step detector-tracker ML pipeline is used in the MediaPipe model. First, the pipeline locates the pose region-of-interest (ROI) within the frame with a detector. Then all 33 pose key points from this ROI are predicted by the tracker. For real-time tracking, only the first frame is run by the detector. Subsequent frames' ROIs are derived and aligned from the previous frame's pose key points. The model's quality has been evaluated against other well-known models. The data shows it has comparable or exceeding performance.

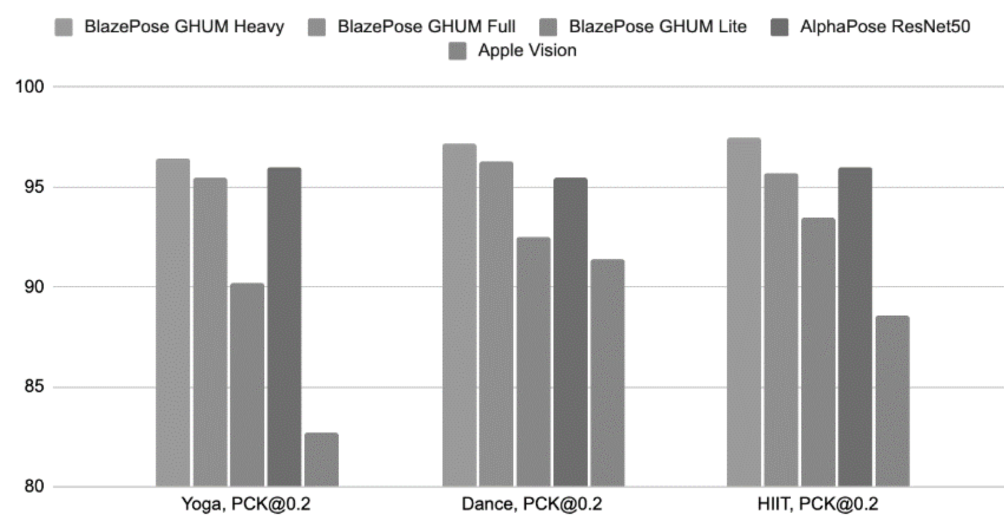


Figure 3. MediaPipe keypoints representation of a human body (MediaPipe, 2021)

Our continuous human pose estimation process involves extracting joint points on a human body and analyzing a human pose using mathematical algorithms. The video recordings or live streams are used as data sources. Keypoints are detected from a sequence of frames, not a static picture. More accurate results are obtained as the model examines a continuous movement of a training session, not a single pose. Our human pose estimation model is implemented in PyCharm (PyCharm, 2021) and MediaPipe is imported into the python interpreter.

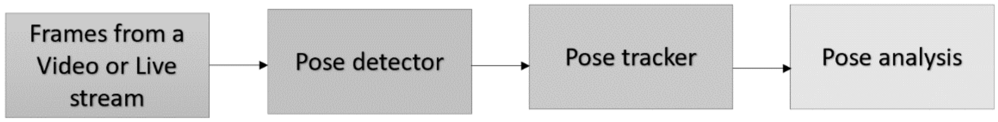


Figure 4. Our human pose estimation model

Tkinter GUI and OpenCV are used to display video frames and build the detection model.

The model includes the following functions:

- Initiate_model: The pose detection model is initiated.
- findPose: Joint key points are detected and the landmarks can be drawn on the body if selected.
- extractLandMarks: Landmarks can be extracted from the body and pulled on the white background.
- findAngle: Angles of arms or legs can be calculated.
- drawCount: The number of exercised repeating postures can be detected and counted.

4. Results

The described pose estimation system is easy to implement on a CPU desktop and efficiently detects and analyzes different fitness movements. From the analysis results, mistakes and corrections can be proposed to the user for coaching purposes.

The application input interface (Figure 5) allows the user to select a video recording or webcam. In addition, a snapshot can be taken to capture an image of pose analysis.

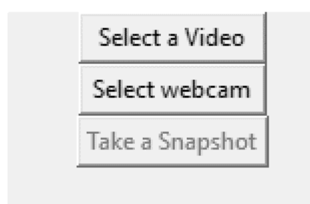


Figure 5. The application input interface

Figures 6 and 7 show examples of the outputs. The left-side panel shows the source frames and the detected body's skeleton. Keypoints are extracted and drawn on the right-side panel. The angles of the arms and knees are calculated. The number of exercises is counted and shown.

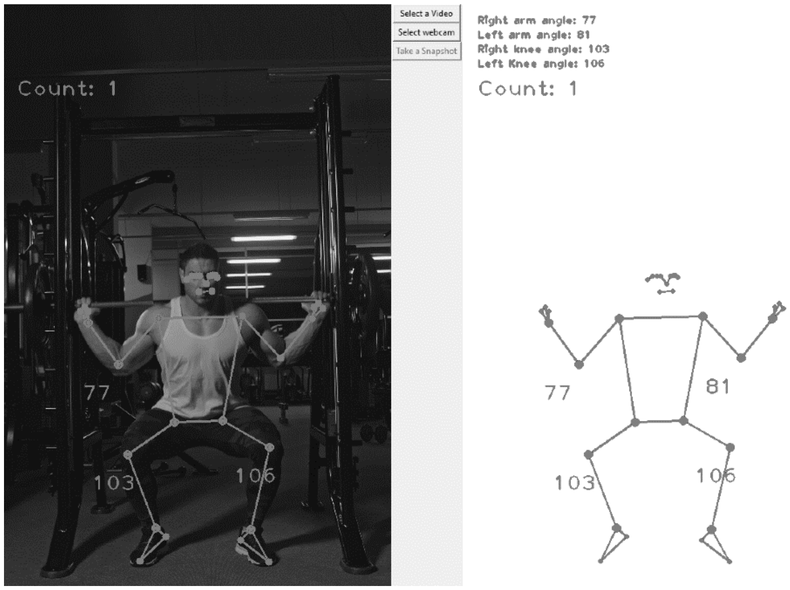


Figure 6. The example of lift pose tracking

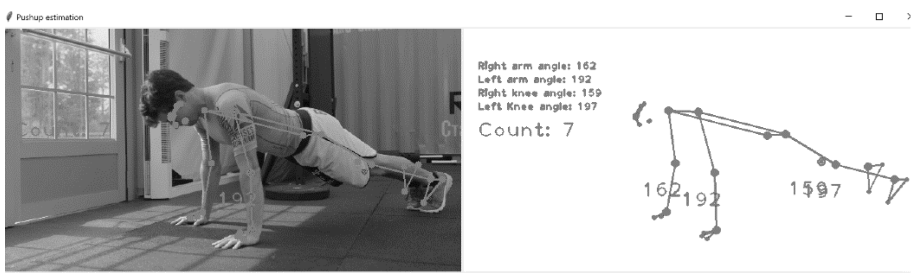


Figure 7. The example of push-up pose tracking

5. Conclusion and future work

In this paper, a demo of a continuous human pose estimation system is presented. A skeleton-based model identifies and tracks the human body joints. The number of the human body's movements and angles of arms and legs are then calculated. The method is straightforward to implement and can run in real-time on a CPU that does not require sophisticated GPU or TPU hardware. For future development, each frame's key points can be compared with a selected reference video's positions to notify the user of the change. That may involve calibration of the distance of the camera. Otherwise, the two videos cannot be compared precisely.

6. References

Pycharm. (2021). Retrieved December 5, 2021, from https://www.jetbrains.com/pycharm/

Digital Synopsis. (2021). Retrieved December 5, 2021, from https://digitalsynopsis.com/design/movies-before-after-green-screen-cgi/

Badiola-Bengoa, A., & Mendez-Zorrilla, A. (2021). A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise. *Sensors*. 21(18):5996. https://doi.org/10.3390/s21185996

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose tracking. *Computer Vision and Pattern Recognition*. https://doi.org/10.48550/arXiv.2006.10204

Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172-186. http://doi.org/10.1109/TPAMI.2019.2929257

Charles, J., Pfister, T., Everingham, M., Zisserman, A. (2014). Automatic and Efficient Human Pose Estimation for Sign Language Videos. *International Journal of Computer Vision*, 110, 70-90. https://doi.org/10.1007/s11263-013-0672-6

COCO 2020 Keypoint Detection Task. (2021). Retrieved December 2, 2021, from https://cocodataset.org/#keypoints-2020

Fang, H., Xie, S., Tai, Y., Lu, C. (2017). RMPE: Regional Multi-person Pose Estimation. 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2353-2362. https://doi.org/10.1109/ICCV.2017.256

Gerónimo, D., López, A., Sappa, A. D., Graf, T. (2010). Survey of Pedestrian Detection for Advanced Driver Assistance Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7), 1239-1258. https://doi.org/10.1109/TPAMI.2009.122

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2018). Mask R-CNN. *IEEE Transactions on Pattern Analaysis and Machine Intelligence*. 42(2):386-397. https://doi.org/10.1109/TPAMI.2018.2844175

Hu, Y., Hugonot, J., Fua, P., Salzmann, M. (2019). Segmentation-driven 6D Object Pose Estimation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3380-3389. https://doi.org/10.1109/CVPR.2019.00350

Luo, H., Wang, M., Wong, P. K.-Y., Cheng, J. C. (2020). Full body pose estimation of construction equipment using computer vision and deep learning techniques. *Automation in Construction*, 110, 103016. https://doi.org/10.1016/j.autcon.2019.103016

Marchand, E., Uchiyama, H., Spindler, F. (2016). Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, 22(12), 2633-2651. https://doi.org/10.1109/TVCG.2015.2513408

MediaPipe. (2021). Retrieved December 3, 2021, from MediaPipe: https://google.github.io/mediapipe/

Panteleris, P., Michel, D., Argyros, A. (2021). Toward Augmented Reality in Museums: Evaluation of Design Choices for 3D Object Pose Estimation. *Frontiers in Virtual Reality*. https://doi.org/10.3389/frvir.2021.649784

Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P., Schiele, B. (2016). DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.1511.06645

Pose Detection. (2020). Retrieved December 3, 2021, from ML Kit: https://developers.google.com/ml-kit/vision/pose-detection

Simon, S. (2004). Quantification of human motion: gait analysis-benefits and limitations to its application to clinical problems. *Journal of Biomechanics*, 37(12), 1869-1880. https://doi.org/10.1016/j.jbiomech.2004.02.047

Zhang, Y., Chen, K., Yi, J., Liu, L. (2014). Pose estimation in physical human-machine interactions with application to bicycle riding. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3333-3338. https://doi.org/10.1109/IROS.2014.6943026

Integration of Instance-based Learning and Computed Torque Control for an Effective Assist-as-Needed Support in Human-Exoskeleton Interaction

Pablo Delgado[1*]
Yimesker Yihun[1]
Lisa Vangsness[2]
Hongsheng He[1]
[1]*College of Engineering, Wichita State University, USA*
[2]*Fairmount College of Liberal Arts and Science, Wichita State University, USA*
padelgado@shockers.wichita.edu

Abstract

The physiological responses that arise from human-robot interaction may vary across subjects in magnitude and rate. Such individual variations may require instance-based learning over model-based learning algorithms. For instance, a wearable assist-as-needed exoskeleton may require real-time progress data to provide the appropriate level of support to a specific user. In this study, an instance-based learning algorithm was developed and integrated with a computed torque control law. Real-time bio-signals, in the form of electromyography (EMG), were tracked during a predetermined time window to quantify an adaptive threshold value and to control the torque at the exoskeleton joints. These signals were fed to the algorithm, which instantly learned and determined the support needed to accomplish a desired task. The algorithm was tested on a 5-degree-of-freedom wearable exoskeleton used in the automation of upper-limb therapeutic exercises. Results indicated that the algorithm offered the ability to adjust assist-as-needed support instantly based on the amount of muscle engagement present in the combined motion of the human-exoskeleton system.

Keywords: Assist-as-Needed, Human-Exoskeleton, Control Systems, Nonlinear Systems

1. Introduction

Wearable exoskeletons are used for augmentation (Farris et al., 2022) and rehabilitation exercises (Zheng et al., 2022). In the former case, these devices are used to augment workers' limbs to help them properly do their jobs without risking injury. For example, a wearable exoskeleton may help a worker carry heavy loads in their day-to-day activities and may also help to automate rehabilitation routines. In both applications, proper interaction and synergy must be established between the human body and the exoskeleton system.

Different algorithms have been developed to this end; One approach is the impedance control

Submitted April 26, 2022
Revised August 22, 2022

strategy, which models the relationship between a desired position and force using a damper-spring-mass system (e.g., (Yang et al., 2006). Another approach involves learning through a neural-network-based control model using offline subject-specific required torque data, then using it to render a force field to assist motion by following a target trajectory (Agarwal & Deshpande, 2017). A third method uses offline EMG signals to estimate joint torques and model predictive control to derive the robot joint torques (e.g., (Teramae et al., 2017)). All three methods are time consuming since a learning step must be performed before using the wearable exoskeleton. Also, the performance of these algorithms will be affected by the varying nature of EMG signals. Therefore, it is preferred to have an instance-based learning assist-as-needed algorithm to update the control law. In this work, we present a hybrid instance-based learning computed torque algorithm to update the assistance provided by the wearable exoskeleton. The proposed algorithm is tested on a 5-degree-of-freedom articulated system.

2. Description of exoskeleton

The exoskeleton system, Figure 1, possesses 5 revolute joints; 4 revolute joints that mimic shoulder kinematics, and 1 that emulates the elbow joint. The dynamics needed to develop and implement the controller can be derived by using either the Euler-Lagrange method (Zefran & Bullo, 2005) or the Recursive Newton-Euler Algorithm (RNEA), based on Newton's laws (Khalil, 2011). The former method uses the kinetic and potential energy of the robot to obtain the equations of motion, yielding a simpler and more elegant set of equations. However, as the number of degrees of freedom (DOF) increases, the complexity of the procedure does as well, making it unfeasible. On the other hand, the RNEA formulation allows the calculation of each link twist (Vector of linear and angular velocities) and wrench (vector of forces and moments) at a given time without the need for derivation and is computationally more efficient to implement for robots with many DOFs. Independently of which method is used, the following closed-form dynamic equation can be derived:

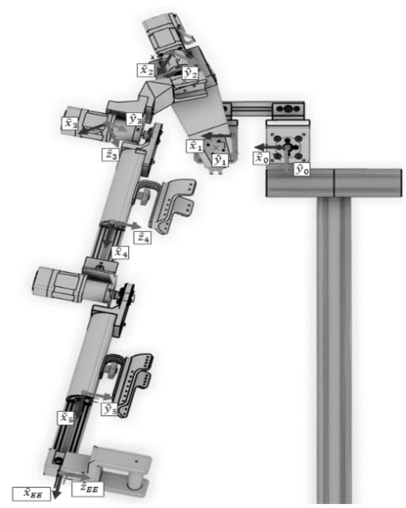


Figure 1. Five-Joint Exoskeleton

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + J(q)^T\mathcal{F} = \tau \tag{1}$$

Where $q \in \mathbb{R}^n$ is a generalized vector of joint coordinates, $M(q) \in \mathbb{R}^{n\times n}$ is the mass matrix, $C(q,\dot{q}) \in \mathbb{R}^{n\times n}$ is the Coriolis & centripetal matrix, $G(q) \in \mathbb{R}^n$ is wrench due to gravity, $J(q) \in \mathbb{R}^{n\times n}$ is the Jacobian matrix, $\mathcal{F} \in \mathbb{R}^n$ is the wrench of forces and moments apply on the environment from the robot at the end-effector coordinates, $\tau \in \mathbb{R}^n$ is a vector of torque and forces exerted by each joint, and n is the number of DOF. Sometimes, it is more convenient to express the summation of $C(q,\dot{q})\dot{q} + G(q) + J(q)\mathcal{F}\ as\ h(q,\dot{q})$, then Equation 1 becomes:

$$M(q)\ddot{q} + h(q,\dot{q}) = \tau \tag{2}$$

In this study, the RNEA is implemented based on forward and backward iteration. In the forward iteration, given a robot with L_{n+1} attached frames from the base to the end-effector, the twist of each link starting from the base frame (L_0) is calculated using Newton's laws and the actual joint positions and rates. Then, using the calculated twists in the backward iteration of the algorithm, the moments and forces exhorted on each link are calculated going backward from the L_{n+1} frame to the L_0 frame. The actual configuration of the robot, the inertia matrices of each link, the link masses, and the center of gravity are required. These parameters need to be defined with respect to the local coordinate frame where each twist is calculated. As shown in Figure 1, each link frame is defined at the center of mass (see (Lynch et al., 2017) for further details).

2.1. Inertia and kinematic parameters of the exoskeleton

The inertia parameters of the exoskeleton are presented in Table 1. Each parameter was estimated from CAD software at the center of mass of each link. In our work, the velocities and forces are derived in spatial coordinates; therefore, the screw axis of each link is derived as well as the initial configuration of the end effector when the joint parameters are set to zero (see Table 2 and Table 3 respectively).

Table 1. Inertia parameters expressed in kg and $kg \cdot m^2 \times 10^{-6}$

Link	1	2	3	4	5
Mass	3.42	0.73	0.73	2.07	1.26
I_{xx}	9,441	1,912	566.5	2,576	595.2
I_{yy}	13,791	639.4	1,720	30,634	9,260
I_{zz}	11,492	1,880	1,533	29,277	9,357
I_{xy}	-1,985	513.7	286.2	-1,618	-719.5
I_{xz}	1,717	1.550	59.34	4,594	354.0
I_{yz}	412.4	-42.91	4.500	353.6	95.58

Table 3. Screw axis of each link

Screw	S_1	S_2	S_3	S_4	S_5
S_x	0.0000	0.0000	-0.3039	-0.9740	-0.9740
S_y	0.0000	0.0000	0.8110	0.2209	0.2209
S_z	1.0000	1.0000	0.5000	0.0498	0.0498
S_{ox}	-0.0223	-0.1740	-0.1704	-0.0476	-0.0565
S_{oy}	-0.0832	-0.1737	-0.1927	-0.2528	-0.3631
S_{oz}	0.0000	0.0000	0.2090	-0.1909	0.5052

Table 2. End-effector initial configuration

$\hat{x}_{ee}$	$\hat{y}_{ee}$	$\hat{z}_{ee}$	$\hat{P}_{ee}$
0.2230	0.0396	-0.9740	0.3454
0.8972	0.3823	0.2209	0.5910
0.3811	-0.9232	0.0498	0.4059
0.0000	0.0000	0.0000	1.0000

3. Computed control law

The robot's equation of motion yields a highly nonlinear system of differential equations that must be linearized to apply linear control system strategies. In the literature, one strategy used to linearize a robot's dynamics is the use of feedback linearization or computed torque control (CTC) (Lynch et al., 2017), (Kelly et al., 2005), (Ullah et al., 2014). The CTC is calculated as follows:

$$\tau_{ctc} = \widehat{M}(q)\left[\ddot{q}_d + k_d\dot{e} + k_p e\right] + \hat{h}(q,\dot{q}) \qquad (3)$$

Where $\widehat{M}(q)$ and $\hat{h}(q,\dot{q})$ are the best estimates of the mass matrix, and the vector of forces and torques due to Coriolis & centripetal forces, gravity forces, and end-effector forces. e is the error between the desired joint parameter vector q_d and the actual joint parameter q ($e = q_d - q$). Substituting the computed torque, Equation (3), into the equation of motion of the robot, Equation (2), yields the following equation:

$$\widehat{M}(q)\left[\ddot{q}_d + k_d\dot{e} + k_p e\right] + \hat{h}(q,\dot{q}) - M(q)\ddot{q} - h(q,\dot{q}) = 0 \qquad (4)$$

Assuming that $\widehat{M}(q) = M(q)$ and $\hat{h}(q,\dot{q}) = h(q,\dot{q})$ yields the error dynamics set of equations:

$$\ddot{e} + k_d\dot{e} + k_p e = 0 \qquad (5)$$

The error dynamics of the system, Equation (5), approaches zero as long as the gain matrices, k_d, and k_p, are positive-definite. Therefore, assuring the stability of the system.

4. Hybrid instance-based learning computed torque algorithm

This study proposes instance-based learning to update an adaptive modifying factor, α, based on current surface electromyography (EMG) data from a subject. This factor calibrates the amount of torque provided by the exoskeleton joints during rehabilitation exercises. Usually, exoskeletons are programmed with given routines – such as elbow flexion – and carry the subject's limb during such exercise. One drawback of such an approach is that the wearer is not motivated to provide any effort to complete the exercise, reducing the efficacy of the rehabilitation session. Incorporating feedback from the wearer into the exoskeleton supplying torque can vary the assistance and improve the synergy of the exoskeleton-human system.

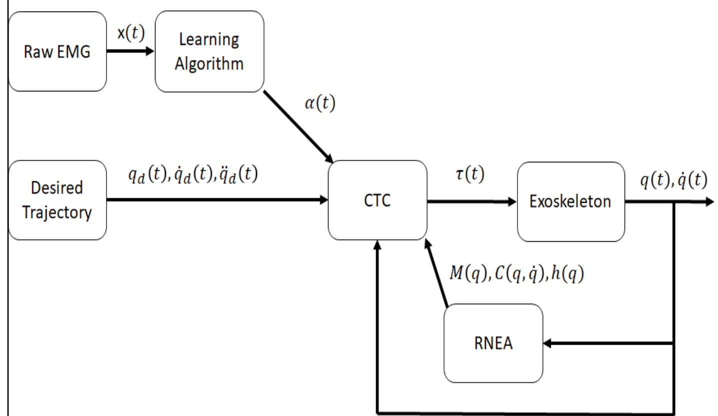
There is evidence that there is a linear relationship between the root mean square (RMS) value of the surface EMG and the contraction force of muscles in the function of the load (Fukuda et al., 2010) (Ullah et al., 2014). The proposed learning algorithm uses a moving RMS window of raw surface EMG data to update the modifying factor (α), then this value is fed back into the control law to modify the calculated torque by CTC in Equation (3). The conditions to update the modifying factor are as follows: First, the maximum RMS value of a given moving RMS time window is extracted, β. Then, this value is compared with new samples, ζ, to determine the value of α, which can be either 1 or ζ/β depending on whether $\beta < \zeta$. This algorithm is shown below along with the proposed closed-loop system. The proposed instance-based learning algorithm is presented in Figure 2a. Depending on the exoskeleton number of DOF, n, there would be *nth* modifying factors, one per each active joint. Grouping these terms together into a modifying diagonal factor matrix, Γ, the desired control torque/force is described in Equation (6).

$$\Gamma = \begin{bmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_n \end{bmatrix}$$

$$\tau_d = (I - \Gamma) \times \tau_{ctc} \qquad (6)$$

```
Required: Raw EMG data
Results: Updated Modifying Factor
Initialization:
t = time window
β= max (movRMS(raw EMG, t))
While (1):
  if new data is available:
    ζ = max (movRMS(raw EMG, t))
    if β < ζ :
      β = ζ
      α = 1
    else:
      α = ζ/β
    end
  end
  if is training done:
    break
  end
```

(a)



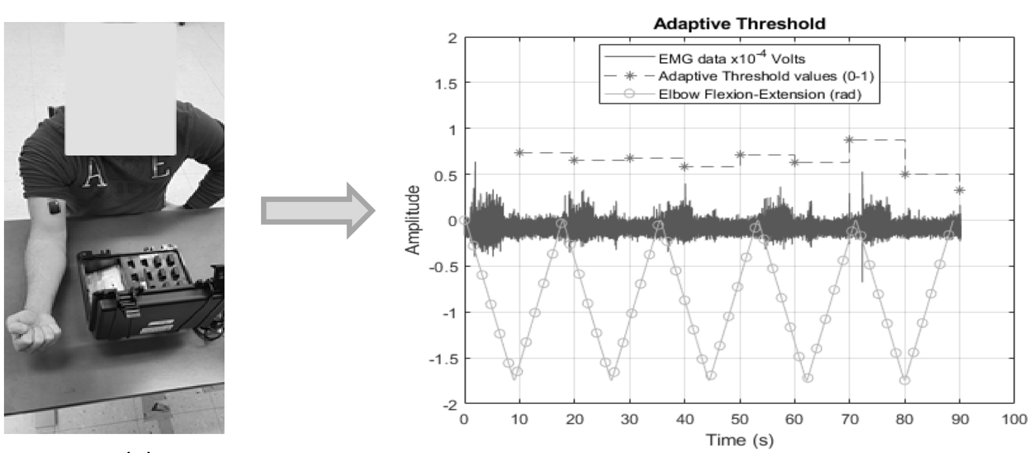
(b)

Figure 2. Proposed Instance-based Algorithm (a), Proposed Closed-loop system (b)

The proposed closed-loop control system is shown in Figure 2b. Where the desired trajectories are fed into the CTC algorithm as well as the learning factor of the targeted muscle, $\alpha(t)$. Also, the system parameters are estimated by the RNEA.

5. Results and discussions

The proposed assist-as-needed algorithm was implemented in an elbow flexion task. The surface EMG signal of the bicep brachii muscle was collected using the Delsys Trigno Wireless EMG system at an approximately sampling rate of $2{,}000Hz$, Figure 3a, when the elbow was flexed from 0° to 100°. Then, the signal was normalized to have zero mean to remove artifacts due to motion (Konrad, 2005), and filtered with a moving RMS window of 0.125 seconds. The task lasted approximately 90 seconds, then the modifying factor was calculated with an updating rate of 10 seconds, Figure 3b. This information was used in MATLAB where the assist-as-needed algorithm was tested along with the CTC.

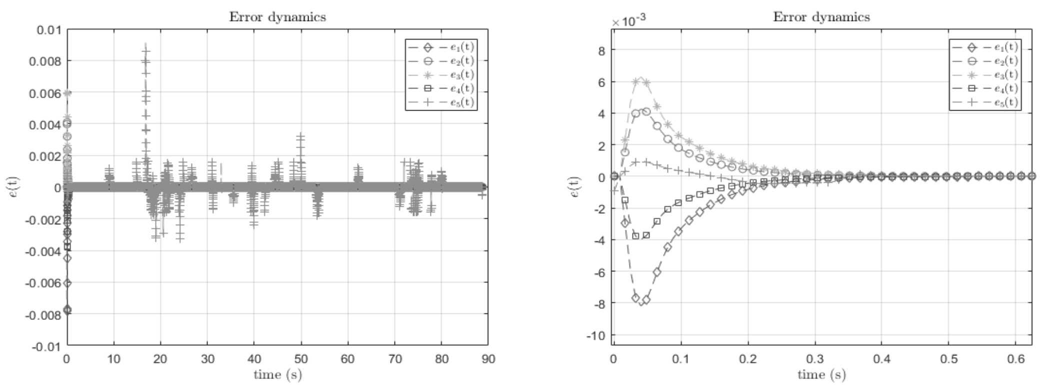


(a) (b)

Figure 3. Data collection system setup (a), Data processing, and learning algorithm (b).

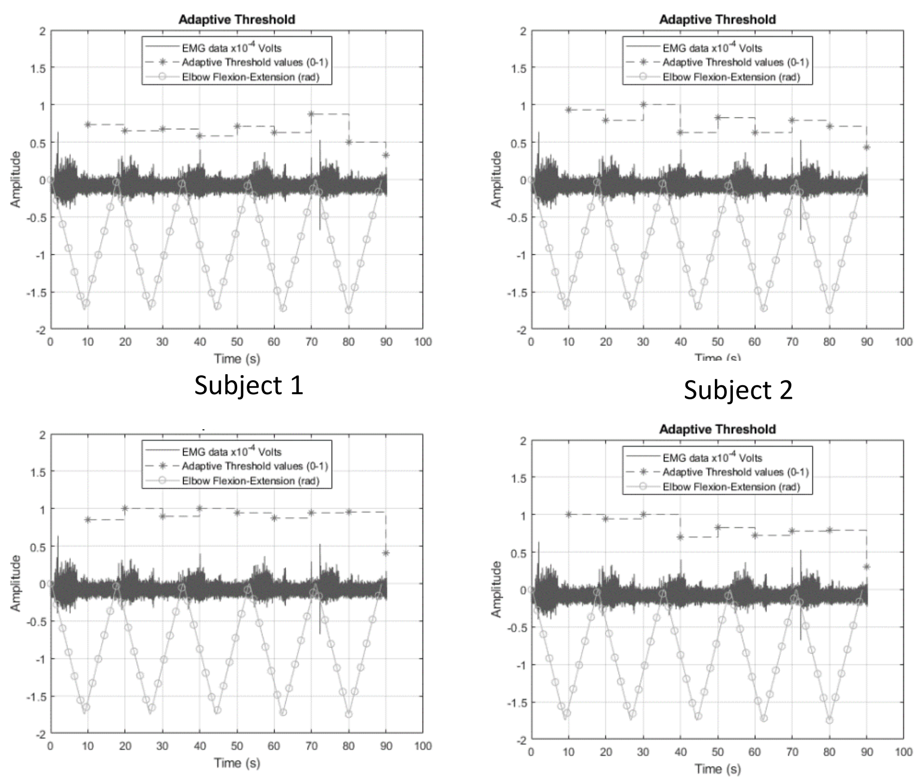
5.1. Implementation without assistance

To demonstrate the effect of adding the EMG instance-based algorithm, the CTC was implemented first to demonstrate that the implemented control law was stable. For this purpose, the gain matrices k_d and k_p were set to $50 \times I^{5\times5}$ and $500 \times I^{5\times5}$, respectably. The error dynamics of the system are shown in Figure 4, as can be inferred from the plot, the system was able to reduce the error dynamics to zero in less than 0.6 seconds.



(a) (b)

Figure 4. Error Dynamics during a 90-second task (a), Error dynamics during the first 0.6 sec (b)



Subject 1 Subject 2

Subject 3 Subject 4

Figure 5. Adaptive Threshold values for each subject

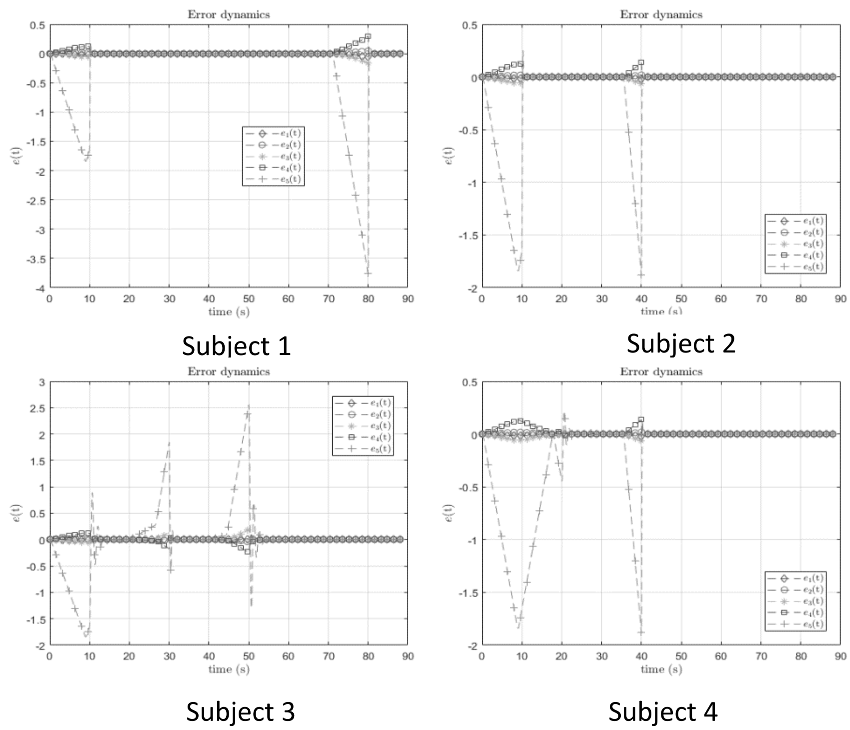


Figure 6. Error dynamics for each subject simulation

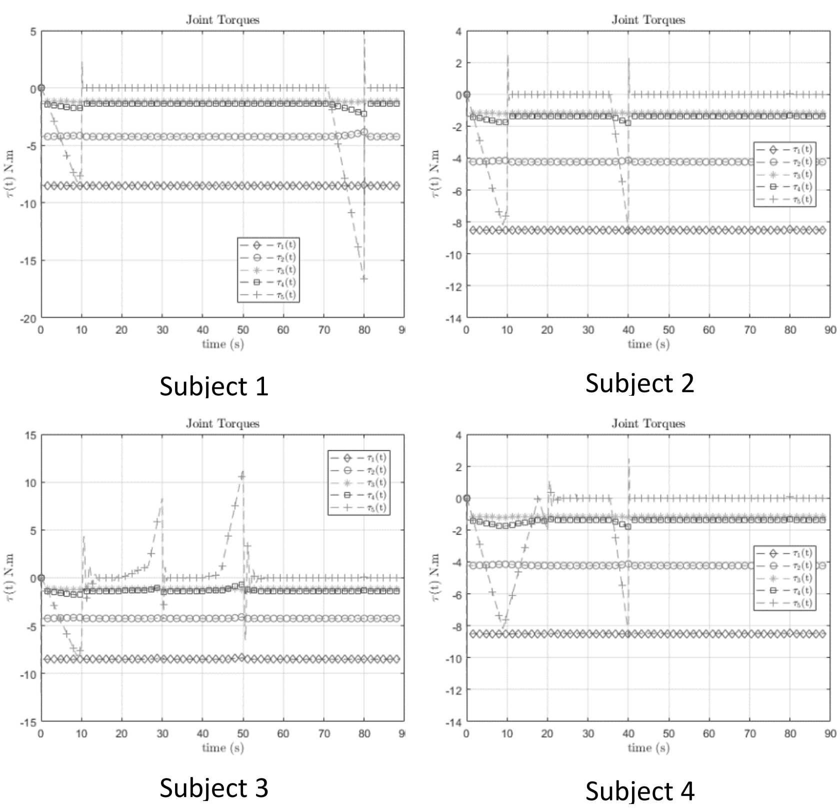


Figure 7. Exoskeleton Joint Torques for each Individual Subject

5.1. Implementation with assistance

Four healthy subjects ages between 18 and 25 volunteered for this study. EMG signals were collected offline for each subject following the above procedure. Modifying factors were calculated, Figure 5, from each subject while performing the elbow flexion task in MATLAB. The error dynamics of the system during each simulation are shown in Figure 6, as well as the joint torque values in Figure 7.

6. Conclusion

In this paper, adaptive EMG-based learning computed torque control algorithm was presented to provide an assist-as-needed control scheme strategy. The proposed method was implemented in MATLAB with data from four different healthy subjects. The assist-as-needed control strategy presented exhibited a dynamic behavior which was expected since the assistance required by each subject varied with the RMS value of the targeted muscle, the bicep brachii. The CTC strategy without assistance could reduce the error dynamics of the system to zero in less than 1 second. On the other hand, when the assist-as-needed control was implemented, the system was capable of correcting the error presented in the desired trajectory for the elbow flexion task in less than 10 seconds for the chosen values of gain matrices k_d and k_p. The error dynamics of the system are expected to be less for real implementation since the simulation does not account for the amount of torque provided by subjects at the moment of the exercise. Despite this, the required torqued from the motors are bounded between reasonable values, $-13N.m$ and $2N.m$.

7. Acknowledgment

This work was supported by WSU Regional Institute on Aging (RIA) and NSF CMMI 2129113 grants.

8. References

Agarwal, P., & Deshpande, A. D. (2017). Subject-specific assist-as-needed controllers for a hand exoskeleton for rehabilitation. *IEEE Robotics and Automation Letters*, *3*(1), 508–515.

Farris, D. J., Harris, D. J., Rice, H. M., Campbell, J., Weare, A., Risius, D., Armstrong, N., & Rayson, M. P. (2022). A systematic literature review of evidence for the use of assistive exoskeletons in defence and security use cases. *Ergonomics*, *just-accepted*, 1–31.

Fukuda, T. Y., Echeimberg, J. O., Pompeu, J. E., Lucareli, P. R. G., Garbelotti, S., Gimenes, R. O., & Apolinário, A. (2010). Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female subjects. *J Appl Res*, *10*(1), 32–39.

Kelly, R., Davila, V. S., & Loria, A. (2005). Computed-torque control and computed-torque+ control. *Control of Robot Manipulators in Joint Space*, 227–241.

Khalil, W. (2011). Dynamic modeling of robots using newton-euler formulation. In *Informatics in Control, Automation and Robotics* (pp. 3–20). Springer.

Konrad, P. (2005). The abc of emg. *A Practical Introduction to Kinesiological Electromyography*, *1*(2005), 30–35.

Lynch, K. M., Park, F. C., Lynch, K. M., & Park, F. C. (2017). *Modern Robotics: Mechanics, Planning and Control* (1st ed.). Cambridge University Press.

Teramae, T., Noda, T., & Morimoto, J. (2017). EMG-based model predictive control for physical human--robot interaction: Application for assist-as-needed control. *IEEE Robotics and Automation Letters*, *3*(1), 210–217.

Ullah, M. I., Ajwad, S. A., Islam, R. U., Iqbal, U., & Iqbal, J. (2014). Modeling and computed torque control of a 6 degree of freedom robotic arm. *2014 International Conference on Robotics and Emerging Allied Technologies in Engineering, ICREATE 2014 - Proceedings*, 133–138. https://doi.org/10.1109/iCREATE.2014.6828353

Yang, Y., Wang, L., Tong, J., & Zhang, L. (2006). Arm rehabilitation robot impedance control and experimentation. *2006 IEEE International Conference on Robotics and Biomimetics*, 914–918.

Zefran, M., & Bullo, F. (2005). Lagrangian dynamics. *Robotics and Automation Handbook*, 1–5.

Zheng, L., Hawke, A. L., & Evans, K. (2022). Critical review on applications and roles of exoskeletons in patient handling. *International Journal of Industrial Ergonomics*, *89*, 103290.

Key Barriers to Industry 4.0 in Construction Industry

Abdelhakim A. Al Turk*
Gamal S. Weheba
Wichita State University, USA
axalturk@shockers.wichita.edu

Abstract

The construction industry makes significant contributions to the national GDP. The rapid increase in urban population resulted in increased demand for quality and productivity. Industry 4.0 offers an opportunity for the construction industry to grow sustainably. The term Industry 4.0 includes a range of technologies used to develop a digital value chain and enable automated manufacturing. The primary objective of this paper is to determine the status of industry 4.0 and its implementation in the construction industry. Using content analysis methods, this study analyzed 108 peer-reviewed articles published from January 2016 to December 2021, to reveal the most important key barriers to a successful implementation of Industry 4.0 technologies in the construction sector. The results shed light on some real challenges that can affect industry 4.0 applications and suggest directions for further research.

Keywords: Industry 4.0, Construction 4.0, Qualitative Analysis, Factor Analysis.

1 Introduction

Industry 4.0 concept started in Germany in 2011 as a vision for advanced manufacturing. As reported by Zhou et al (2015), to solve the European debt crisis and as a recovery plan, the German Government presented a strategy for Industry 4.0, to consolidate and promote global German manufacturing power. The main aim of Industry 4.0 is to increase the digitization of production. Industry 4.0 is the fourth industrial revolution, which is led by intelligent manufacturing. The fourth revolution went beyond the industry, and it reaches the construction sector, it is tagged as "construction 4.0". Therefore, Construction 4.0 represents all technological changes related to the implementation of new work methods which are related to processes, materials, and markets. In other words, according to Sawhney et al (2020), construction 4.0 is the adoption of the industry 4.0 framework and technologies for the construction sector. These technologies include 3d printing, automation, 3d scanning, virtual reality, drones, reverse engineering, sensors, artificial intelligence, robots, data mining, and more.

The following section represents a literature review of the implementation of industry 4.0 in the construction industry. The third section illustrates the research methodology then the data and results are discussed in the fourth section. Final conclusions, research limitations, and future research are highlighted in the fifth section.

2 Literature Review

As stated by Brettel et al (2017), industry 4.0 is a new industrial stage in which there is an integration between manufacturing operations systems, information, communication technologies,

Submitted April 29, 2022
Revised August 10, 2022

the internet of things, and cyber-physical systems. In other words, industry 4.0 seeks to describe the intelligent factory, with all processes interconnected by the Internet of Things. This integration of industry 4.0 has a lot of benefits to the employees and business. Nowadays, employees have more power than ever in their relationships with employers. Industry 4.0 supports employee growth by creating a management culture that encourages communication and training. As reported by Kagermann et al (2015), industry 4.0 allows employees to advance their skills. Employees will be able to be trained and embrace the process of continuous learning. The enforcement of industry 4.0 has a lot of benefits for the company and business. It can reduce product costs, increase productivity, lower operating costs, and improve product quality. According to Telukdarie et al (2018), virtual planning systems, simulation models, forecasting, analysis, and synthesis, all virtually help production and reduce manufacturing costs. As stated by Vaidya et al (2018); industry 4.0 allows manufacturers to visualize and analyze real-time data, which leads to production optimization and reinforcing economic competitiveness. In addition to that, Rüßmann et al (2015), reported that industry 4.0 helps employers to adopt new business models, production processes, and other innovations. This will lead to a new level of mass customization production with lower cost and higher quality. It allows for a faster response to customer needs. Moreover, it reduces production costs and improves the speed, productivity, flexibility, and quality of the product and the production process.

The prospects of the global construction industry are promising. The main driving forces of the market are the growth of the construction activities like urbanization, the housing market, and infrastructure. According to Gurney et al (2020), the construction industry will be an engine of global economic growth in 2030, with output expected to be 35% higher than in the ten years to 2020, according to a new global forecast. The fund spent on the construction industry accounted for 13% of global GDP in 2020 and it is expected to reach over 13.5% in 2030. A cumulative total of US$135 trillion in construction output is forecasted in the decade to 2030.

With the growth of the construction industry and the technological advancements in construction technology, the construction 4.0 market has experienced significant progress over the past decade. As stated by García et al (2019), by adopting construction 4.0, the construction industry will no longer be fragmented but will become a highly integrated industry, with an integrated construction process, organization, and projects. As reported by Forcael et al (2020), construction 4.0 derives from the foundation of industry 4.0 but focuses on and relates to the construction sector.

3 Research Methodology

The primary aim of this research was to identify the key barriers to implementing industry 4.0 in the construction industry. NVivo (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12, 2021) is a software program used for qualitative and mixed-methods research. In this research, NVivo software was used to analyze the textual contents of 108 articles in the English language in peer-reviewed journals published between January 2016 and December 2021. These 108 articles are related to the implementation of industry 4.0 in the construction industry. Computer text analysis was chosen for its ability to analyze the textual content of a large number of articles. The analysis provided the number of occurrences of each barrier within the selected sample of 108 publications. The research methodology is depicted in Figure 1, and a detailed description of the steps taken is represented in the following sections.

3.1. Initial dimensions

Based on the literature review of the implementation of industry 4.0 in the construction industry, a list of six dimensions was assembled. The list includes barriers (economic, environmental, organizational, security, social, and technological) found in the literature that can affect the implementation of industry 4.0 in the construction industry. These six dimensions and their source are listed in Table 1.

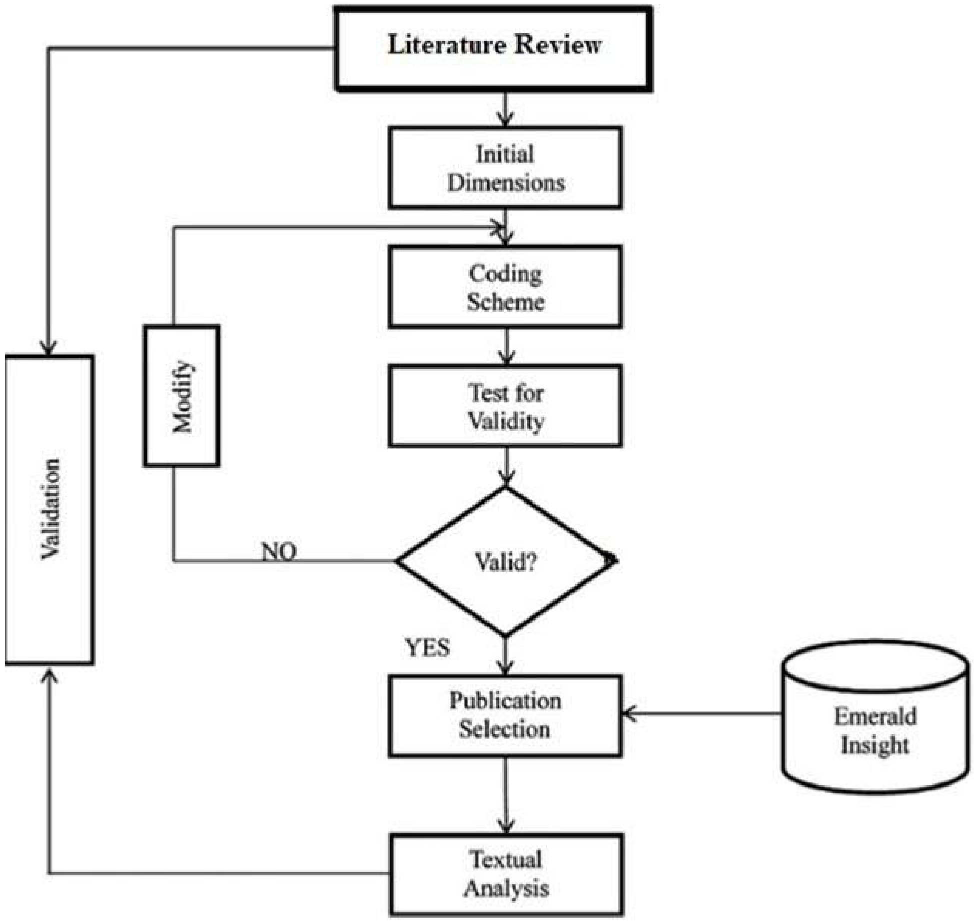


Figure 1. Research Methodology

Table 1 Dimensions & Sources

	Dimension	Source
1	Economic	Michaloski et al (2010), Erdogan et al (2010)
2	Environmental	Son et al (2010), Abubakar et al (2014)
3	Organizational	Zakari et al (2014), Erdogan et al (2010), Popov et al (2010), Wong et al (2014)
4	Security	Zakari et al (2014), Abubakar et al (2014), Wong et al (2014), Mahamadu et al (2013), Volk et al (2014)
5	Social	Michaloski et al (2010), Abubakar et al (2014)
6	Technological	Zakari et al (2014), Abubakar et al (2014), Erdogan et al (2010), Wong et al (2014)

3.2. Coding scheme

A list of possible codes representing each dimension was proposed. The list of codes was validated by a panel of faculty members working in Industry 4.0 at Wichita State University. Through the validation process, faculty members recommended the inclusion of each code or not. They also could recommend additional codes based on their field experience. To avoid bias, each dimension was given seven codes. The list of final codes used to perform the textual analysis is shown in Table 2.

Table 2: List of Final Codes

Dimension	Code
Economic	Investment, Cost, Profit, Gains, ROI, Benefits, Loss
Environment	Energy Consumption, Sustainability, Climate Change, Global warming, Recycling, Fuel, Atmosphere
Organizational	Development, Skill, Transformation, Training, Support, Planning, Suppliers
Security	Legal, Data, Information, Safety, Risks, Policy, Contractual
Social	Knowledge, Resistance, Innovation, Adoption, Behavior, Culture, Company Size
Technological	Implementation, Automation, Computing, Software, Digitization, Connectivity, Programming

3.3. Publication selection

An initial search of publications with titles containing the word 'Industry 4.0' was performed. A second search was performed for publications with tiles containing 'Construction 4.0'. The search included all articles listed in the Wichita State University library database (https://libraries.wichita.edu/home) in the English language. To ensure reporting accuracy, the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines were followed. As shown in Figure 2, the initial search resulted in 2,587 results related to industry 4.0 this number was reduced to 1,556 by limiting the search to peer-reviewed articles only. The remaining number of publications was further reduced from 1,556 to 61 by limiting the search to publications related to the construction industry. These 61 publications were considered for further research. The second search of construction 4.0 resulted in 1,095 publications related to construction 4.0. This number was reduced to reach 47 articles, by limiting the search to peer-reviewed articles only. As a result, 108 publications were considered units for textual analysis. These 108 publications were divided into 61 publications related to industry 4.0, and 47 publications related to construction 4.0.

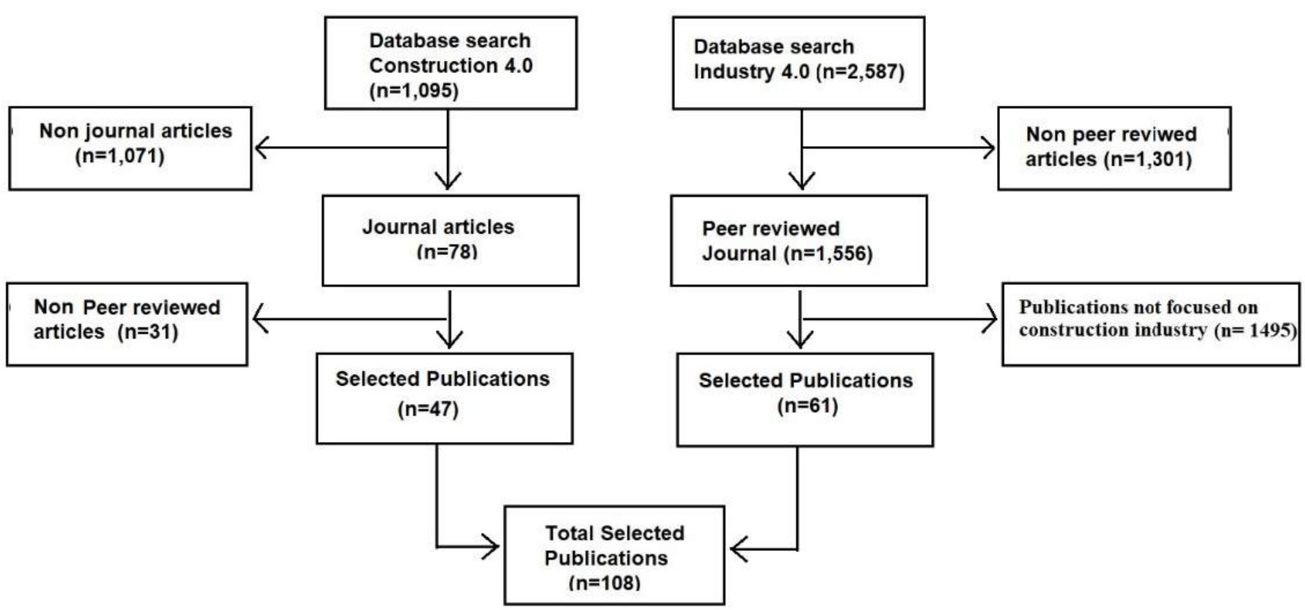


Figure 2: PRISMA Flowchart for Publication Selection

4 Data Analysis

In 1950, Joseph M. Juran rephrased the theories of Italian economist Vilfredo Pareto (1848-1923) as the Pareto principle, often referred to as the 80-20 rule 20% of the sources cause 80 % of the problem. As noted by Sanders (1987), this tool helps in the prioritization of the reasons which contribute to 80% of the problem. The rule postulates that in any series of variables (problems or errors), a small number will account for most of the effect. A Pareto analysis of the results was performed to recognize the vital dimensions contributing to the majority of the count. In other words, the analysis provided the number of occurrences of each dimension and its code within the selected sample of 108 publications. The Pareto chart in Figure 3 indicates that security, social, technological, and organizational represent almost 85% of the total count. It can be noted that these four dimensions represent the most common dimensions used within the sample of publications to identify barriers to implementing industry 4.0 in the construction industry. Also, it does not appear that environmental and economic barriers received enough interest in published research.

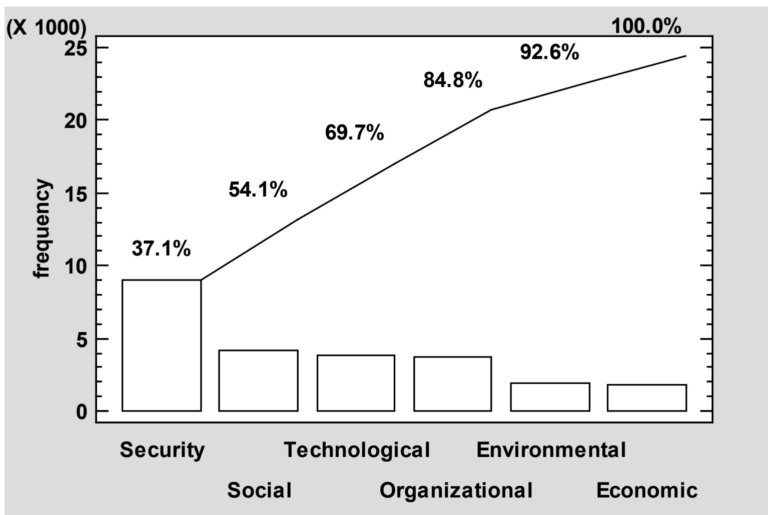


Figure 3. Pareto Chart for Frequency

A contingency table was constructed using the publications as rows and the dimensions as columns, this resulted in a 108 X 6 matrix. Then Statgraphics software (Statpoint Technologies Inc., Centurion version 18,2022) was used to test the analysis. Bartlett's test of Sphericity was used to test the hypothesis that the correlation matrix amongst the variables is an identity matrix, indicating that they share no common variance. Since the test resulted in a Chi-Square = 146.138 and a P-value equal to 0.0001 and less than 0.05, so it can be concluded that the hypothesis is rejected.

As reported by Cureton (2013), factor analysis consists of a group of procedures for analyzing the relations among a set of random variables observed, counted, or measured within the group. Factor analysis deals with the relations among the random variables, it provides a score for each variable. As stated by DeCoster (1998), factor analysis is a collection of methods used to examine how underlying constructs influence the responses to several measured variables.

In addition, the factorability test was used to indicate whether or not it is likely to be worthwhile attempting to extract factors from a set of variables. The Kaiser-Meyer-Olsen KMO measure indicates how much common variance is present. As noted by Saphores et al (2009), for factorization to be worthwhile, KMO should normally be at least 0.6. Since KMO = 0.72, it can be concluded that factorization is likely to provide interesting information about the underlying factors. The purpose of the factor analysis is to obtain a small number of factors that account for most of the variability.

Table 3 illustrates the results of the factor analysis. It shows the Eigenvalue, percentage of variance, and cumulative percentage of each of the total counts. The screen plot suggested that the six dimensions can be presented using two factors, with an eigenvalue greater than 1.0. These two components together account for 62% of the variability in the total count.

Table 3. Results of Factor Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	2.62072	43.679	43.679
2	1.12036	18.673	62.351
3	0.856545	14.276	76.627
4	0.581205	9.687	86.314
5	0.429725	7.162	93.476
6	0.391442	6.524	100.000

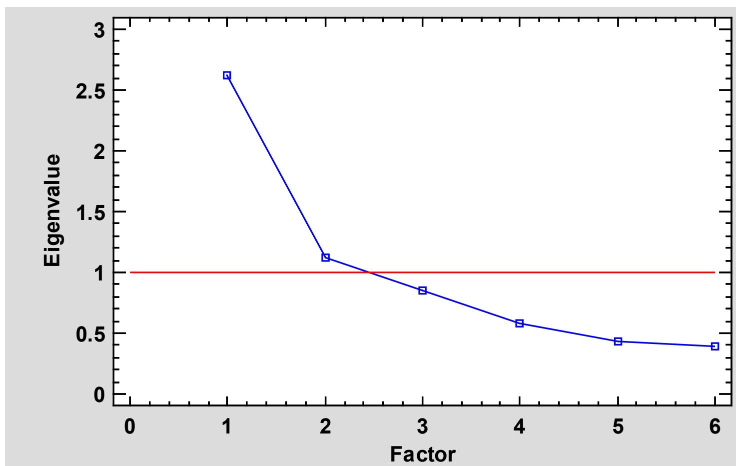


Figure 4. Screen Plot

It is important to mention that the six dimensions considered in this research are not claimed to be comprehensive. A different number of articles or different citations may lead to new dimensions. The results of the factorial analysis are presented in Table 4. These results indicate that the six dimensions can be reduced to two key factors meta dimensions: administrative barriers and security barriers.

Table 4. Factor Score Coefficients

	Factor 1	***Factor 2***
	Administrative	***Security***
Security	0.396645	0.837803
Social	0.79211	-0.2625
Technological	0.721264	0.360523
Organizational	0.684001	-0.349827
Environmental	0.555005	0.119618
Economic	0.734742	-0.287885

Administrative barriers included social, economic, technological, organizational, and environmental dimensions. Social barriers appear to have the highest weight within the first factor (79%). Security barriers contributed 83% to the second factor. These results suggest that security standards are needed to protect corporate data assets and support the implementation of construction 4.0. In addition, social barriers can be attributed to the nature of the construction industry and its business model.

5 Conclusions

Over the last few years, industry 4.0 has been introduced as a popular term for the integration between manufacturing operations systems, information, communication technologies, the internet of things, and cyber-physical systems. The adoption of industry 4.0 in the construction industry will enable the digitization, automation, and integration of the construction processes.

The primary aim of this research was to explore the key barrier to industry 4.0 in the construction industry. The results of this research indicated that published research articles have focused on security, social, technological, and organizational dimensions as the main barriers to implementation. Furthermore, within the sample of publications, it can be concluded that security was the most frequently cited barrier, and economics (consider finances?) was the least frequently cited barrier to implementation. In addition, the results of the factor analysis suggested two main factors, administrative and security, that appear to dominate the key barriers to implementing industry 4.0 in the construction industry.

Finally, it is important to shed the light on some important limitations of this research. The six dimensions considered in this research are not claimed to be comprehensive. Furthermore, time is an important factor in determining the stability of the identified barriers. The results presented in this paper should be viewed as a static representation of barriers addressed in published research (January 2016, December 2021). Research efforts over future periods would prove the longitudinal validity of the proposed dimensions. In addition, in this research 108 peer-reviewed publications were used for the textual analysis, and no attempts were made to stratify the publications based on region or number of citations. Stratifications of publication according to these factors may help identify additional factors.

6 References

Abubakar, M., Ibrahim, Y. M., Kado, D., & Bala, K. (2014). Contractors' perception of the factors affecting Building Information Modelling (BIM) adoption in the Nigerian Construction Industry. In Computing in civil and building engineering (2014) (pp. 167-178).

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2017). How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. *International Journal of information and Communication Engineering*, 8(1), 37-44.

Cureton, E. E., & D'Agostino, R. B. (2013). Factor analysis: An applied approach. Psychology Press.

Erdogan, B., Abbott, C., & Aouad, G. (2010). Construction in year 2030: developing an information technology vision. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924), 3551-3565.

Forcael, E., Ferrari, I., Opazo-Vega, A., & Pulido-Arcas, J. A. (2020). Construction 4.0: A literature review. *Sustainability*, 12(22), 9755.

García de Soto, B., Agustí-Juan, I., Joss, S., & Hunhevicz, J. (2019). Implications of Construction 4.0 to the workforce and organizational structures. *International Journal of Construction Management*, 1-13

Gurney Richard, Simon Liley, Kelly Outram, Amir Mirfenderesky (2020) Future of Construction A Global

Forecast for Construction to 2030.

Juran, J. M. (1950). The role of statistics as a tool of management. *Statistica Neerlandica*, 4(1-2), 69-79.

Kagermann, H. (2015). Change through digitization—Value creation in the age of Industry 4.0. In Management of permanent change, 23-45. Springer Gabler, Wiesbaden.

Mahamadu, A. M., Mahdjoubi, L., & Booth, C. (2013, December). Challenges to BIM-cloud integration: Implication of security issues on secure collaboration. In 2013 IEEE 5th International Conference on Cloud Computing Technology and Science, 2, 209-214.

Michaloski, A. O., & Costa, A. P. C. S. (2010). A survey of IT use by small and medium-sized construction companies in a city in Brazil. *Journal of Information Technology in Construction* (ITcon), 15(28), 369-390.

Popov, V., Juocevicius, V., Migilinskas, D., Ustinovichius, L., & Mikalauskas, S. (2010). The use of a virtual building design and construction model for developing an effective project concept in 5D environment. *Automation in construction*, 19(3), 357-367.

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston Consulting Group*, 9(1), 54-89.

Saphores, J. D. M., Nixon, H., Ogunseitan, O. A., & Shapiro, A. A. (2009). How much e-waste is there in US basements and attics? Results from a national survey. *Journal of Environmental Management*, 90(11), 3322-3331.

Sawhney, A., Riley, M., & Irizarry, J. (2020). Construction 4.0: An Innovation Platform for the Built Environment (1st ed.). Routledge.

Son, H., Kim, C., Kim, H., Han, S. H., & Kim, M. K. (2010). Trend analysis of research and development on automation and robotics technology in the construction industry. *KSCE Journal of Civil Engineering*, 14(2), 131-139.

Telukdarie, A., Buhulaiga, E., Bag, S., Gupta, S., & Luo, Z. (2018). Industry 4.0 implementation for multinationals. *Process Safety and Environmental Protection*, 118, 316-329.

Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0–a glimpse. *Procedia manufacturing*, 20, 233-238.

Volk R., Stengel J. & Schultmann F. (2014). Building Information Models (BIM) for existing buildings – Literature review and future needs, *Automation in Construction*, 38, 109-127.

Wong, J., Wang, X., Li, H., & Chan, G. (2014). A review of cloud-based BIM technology in the construction sector. *Journal of information technology in construction*, 19, 281-291.

Zakari, Z., Ali, N. M. A., Haron, A. T., Ponting, A. M., & Hamid, Z. A. (2014). Exploring the barriers and driving factors in implementing building information modelling (BIM) in the Malaysian construction industry: A preliminary study. *The Journal of the Institution of Engineers*, Malaysia, 75(1).

Zhou, K., Liu, T., & Zhou, L. (2015, August). Industry 4.0: Towards future industrial opportunities and challenges. In 2015 12th International conference on fuzzy systems and knowledge discovery (FSKD), 2147-2152.

7 Appendix

The list of references for the 108 peer-reviewed articles used in the content analysis is available on: https://bit.ly/industry4key

Business Analytics at Florida Commuter Colleges: The Impact and Effectiveness of Implementing a Business Analytics Program

Joachim P. Gooden*
Rakeem R. Ford
Jason T. Black
Florida Agricultural & Mechanical University, USA
joachim1.gooden@famu.edu

Abstract

Only forty of Florida's one-hundred seventy-eight colleges and universities are public, meaning the vast majority are privately run. Many of these schools provide associate degrees or certificates (Community College Review, 2021). Commuter colleges hold the distinction of providing off-campus student living, with the majority offering two-year degrees. Despite the growing need in industry for individuals who can manage large amounts of data, commuter colleges rarely offer business analytics courses, causing many students to miss out on such employment opportunities. Most businesses struggle with analytical planning and are looking for experts who can turn data into insight (Albright & Winston, 2016). The absence of data analytics capabilities causes organizations to waste data resources at a rate of approximately sixty to seventy percent (Forrester, 2017). Intrinsically, universities of all sizes, including commuter colleges, need to train and produce more data analysts. Many major universities now offer business analytics degrees, yet business analytics degrees are available at less than twenty percent of commuter colleges. The goal of this research is to investigate whether commuter colleges would indeed benefit from business analytics programs and to determine the appropriate analytics degree curricula most optimal for these types of institutions. The paper will present an exhaustive comparison of the state of Florida's colleges and universities, including commuter colleges, examining both business intelligence and business analytics degree programs. The objective is to analyze the impact and effectiveness of these programs at larger universities and present a model for developing such programs at commuter colleges.

Keywords: Business Analytics, Big Data, Education Program Effectiveness.

1. Introduction

Many organizations collect massive amounts of data about their competitors, customers, and supply chains (Freedman, 2020). It is easy to obtain data; however, determining its value becomes a persistent challenge. In such a case, the work of business analytic professionals is highly needed (Gavin, 2019). Most businesses encounter problems with planning and as such need experts who can turn their large data into meaningful information for decision-making (Albright & Winston, 2016). As reported by Forrester analysts Mike Gualtieri and Noel Yuhanna, about 60 to 73 percent of the data generated by companies go unused and this can be attributed to the lack of data analysis in these companies to utilize the data (2017). Another report by Paul Leonardi and Noshir Contractor (2018) indicates that more than 70% of companies consider analysts as priorities for their business growth. Intrinsically, there is a need for universities and colleges to produce more data analysts.

However, not all higher learning institutions offer business analytics and related courses. There are 28 colleges and universities in the Florida College System (Fldoe, 2021). In retrospect, there are 12 public

Submitted April 29, 2022
Revised September 6, 2022

four year-year universities in the State University System of Florida (Florida Board Of Governors, 2019). Both of these systems add up to 40 colleges and universities, which make up the Florida system of public higher education. Compared to the vast amount of Business Analytics degree graduates in larger universities, the majority of small colleges don't offer business analytics-related courses. 80% of the 10 largest universities in Florida carry Business Analytics Degrees (Gooden, 2021). In retrospect, out of the 10 commuter colleges the researchers sampled, less than 20% offer business analytics as a degree path (Gooden, 2021). For this research, the designation "commuter college" includes community colleges and small institutions with less than 30k student population. In Florida, these institutions have minimal business analytic programs in contrast to larger institutions such as universities. Should these commuter colleges install the program, they will be able to compete with larger institutions for the higher education business analytics market share. Therefore, in this paper, the researcher will determine the impact and effectiveness of the program in commuter colleges.

1.1 Goal Of research

This paper aims to assess the effectiveness of implementing a business analytics program at a Florida commuter college. The researcher will achieve the objectives of the study by answering the research questions. 1) Do diverse programs like Business Analytics have the possibility of increasing Student Retention at Commuter Colleges? 2) Could a curriculum developed for commuter colleges close the knowledge and skill gap in terms of how Business Analytics is taught among commuter colleges and universities?

1.2 Relevance of research

The need for business analytical skills, which has been consistently growing in the last decade, has been recognized in academia. Presently, there is no model curriculum for business analytics programs at either graduate or undergraduate level. Michael Fitzgerald (2015) emphasized that there is an increasing demand for BA skills even in non-IT fields, making a case for academia's need for a model curriculum (2015). Since available research studies have not assessed the effectiveness of implementing a business analytic program in commuter colleges, this current study will bridge this gap by assessing the effectiveness of implementing a business analytic program in such colleges, in addition to the implementation of business analytics in achieving greater business value through superior insight.

2. Literature review

2.1 Big data/data analytics

Big data and data analytics are currently buzzwords in technology, industry, and academia domains considering the significant increase in the magnitude of data generated by research institutions, businesses, and various organizations in the digital sphere. Big data is defined as a collection of data that is either too voluminous or too structured to be managed and analyzed using traditional methods or tools of data analysis (Davenport, Barth & Bean, 2012). An example of big data used in the industry is clickstream data from the web that companies collect to then send a rotation of targeted ads to their customers (Davenport, Barth & Bean, 2012, p.2). Another case of big data utilized in the industry and also in a Florida University is data science, which bridges IT and management. For example, Florida State University offers a master's degree in data science to equip learners with skills to read, analyze, explore, model, and draw conclusions from the highly complex, multi-dimensional, and diverse data universe (FSU, n.d). One of the most notable scholars that have researched big data is Rakesh Agrawal. A previous study published by Agrawal is entitled "Challenges and Opportunities with Big data." In this study, Agrawal and his colleagues outlined both the challenges and opportunities facing big data. In terms of opportunities, the authors noted that since most big data are generated electronically, there is an opportunity to influence their creation to influence later linkage. The challenges associated with big data as discussed by the authors were that given the large volume of data, there is a challenge in deciding which data to keep and which one to discard. The second challenge is transforming contents into a structured format for later analysis (Agrawal, 2011,p.9). There is also the challenge of data analysis given the complexity of data and lack of scalability. The researchers concluded by encouraging investing in big data stating that investing in big data is laying the foundation for advancement in science, business, and medicine.

Data analytics can be defined as the science of analyzing raw data to make interpretations, and discoveries of meaningful patterns in the data (Moreira, Carvalho & Horváth, 2019, p.4). An example of data analytics is a satellite that contains images and data through surveillance cameras. The data collected through satellites are used to make analytical decisions. An example of the common usage of data analytics in multiple industries and also in Florida universities as a program is supply chain management. Supply chain management uses data analytics to ensure greater insight, clarity, and accuracy in the distribution channel (Ghani et al. 2019). Florida International University offers both a Master of Science degree in Supply Chain Management and a Master's in Business Administration with a concentration in Business Analytics (FIU, n.d). Among the notable researchers who have researched data analytics are Chris Surdak and Sara Agarwal. In one of the studies about how data analytics can be used to drive growth in developing countries, Surdak and Agarwal (2014) noted that big data and data analytics can be used over smartphones to make decisions that could spark development. For example, they indicated that data analytics through smartphones could be used by farmers to make decisions regarding crop production as well as the sale of farm products. They indicated that data analytics could be used not only to answer the question of how but also why. They conclude that mobile apps created by locals in developing countries could be key in unlocking development in those countries.

2.2 Business analytics

The Gartner Group defines business analytics as a tool for creating scenarios, modeling reality, and predicting future events through the use of analytical models and simulations (Gartner, n.d). According to Shabbir and Gardezi (2020), analytics plays a huge role in providing the right tools for making the right decisions and developing strategies necessary to gain a competitive edge in a competitive market and grow profitably. A previous study by MGI research, led by director James Manyika (Manyika, 2011) explored the multiple different methods in business analytics: 1) descriptive, which identifies trends and patterns, 2) diagnostic, which determines why something has happened, 3) predictive, which forecasts future outcomes; and 4) prescriptive, which involves the application of testing and other techniques to determine the outcome that will yield the best result in a given scenario. All the analytics methods are critical for informed data-driven decision-making through data mining.

Sales and marketing is an instance where business analytics is commonly used in multiple industries. Trying to grow corporate sales utilizes business analytics, by entailing the online retail implementation of a sales dashboard (Pappas et al., 2018). A notable scholar in the field of Business analytics is Dursun Delen who together with other colleagues conducted a study to determine the impact of business analytics on business process performance. They were interested in determining the mediating role of business process performance (BPER) on the relationship between business analytics and company performance. They collected data from medium to high-level business executives and conducted an analysis. Their analysis revealed that BPER was a significant influencer in the relationship between business analytics and firm performance (Delen, et al, 2019).

2.3 Business analytics at smaller colleges

In researching the web portals of smaller Florida colleges, most of them do not offer business analytics courses even though most have some related data degrees. Of the 10 small Florida colleges surveyed, 9 colleges did not have any program associated with business analytics; the commuter school with the closest resemblance to a business analytics course was Hillsborough College, located in the west-central portion of Florida (Gooden, 2021). Hillsborough only has a Business Intelligence Specialist Associates Degree pathway (under its IT program) and a Business Intelligence Professional College Credit Certificate (Gooden, 2021). These were offered a four-course sequence program with essential analytic tools, mainly a database and a BI course (Gooden, 2021). Other included courses in these programs focus on related components like Statistics, Programming, and Computer Information Technology Literacy.

A previous study by researchers David Carroll, Kris Ryan, and Susan Elliot titled "*Want to improve your chances of getting a full-time job? A double degree can do that"* depicted how most students combine related degree courses to increase their employment chances (Carroll et al, 2021). The research study concludes by stating that Universities that develop and promote double degrees benefit both students who complete them and the employers they attract (Carroll et al, 2021). Business analytics has a broader focus and covers business intelligence, reporting, and online analytical processing. Offering analytics programs at smaller colleges can promote the production of graduates who combine data analytics skills with different programs of their choice. For example, marketing students in a small commuter college can take a minor in Business Analytics to increase their success in the job market following graduation. Business Analytics programs in smaller colleges may also attract commuting learners from bigger universities interested in multiple courses.

2.4 Summary

To summarize the former sections, the researchers' exposed that most Florida commuter colleges do not offer business analytics courses even though the majority of large Florida universities do offer them. However, the research also revealed that most of the smaller colleges in Florida offer analytics-related courses, including data degree programs, statistics, and information science courses that closely relate to data science (Gooden, 2021). Students pursuing technical routes like Computer Science and Statistics may develop an interest in Business Analytics to increase their chances of getting jobs in their respective domains.

3. Description of research

This research is descriptive, as data was collected from several web portals of Florida educational institutions and processed to characterize them. The variables under study are the following: type of institution (University/College), type of University (Large Institutions/Commuter College), which are qualitative variables with an ordinal level of measurement, Degrees offered, Course line, which are qualitative variables with a nominal level of measurement, and total enrollment from 2017 to 2019, which is a quantitative variable with a ratio level of measurement. The analysis performed for the qualitative variables consisted of the elaboration of frequency distributions, and due to the qualitative nature of the variables, the measure of central tendency that can be presented is the mode of each variable, and pie charts were elaborated to show the percentage distribution of the categories of courses by type of Institution. In addition, stacked bar charts were produced to compare the distribution of the annual enrollment of the selected institutions. For the quantitative variable, the average value and standard deviation are presented.

Research efforts identified which Florida institutions had Business Analytics and related courses as well as if enrollment retention was improving through the 3 years before the pandemic. All of the University and Commuter College data collected were obtained from the Florida Board Of Governors and Florida Department of Education directories (Flbog 2022 and Fldoe 2022). All of the enrollment and course offerings data were compiled from Data USA and Florida learning institution web portals (Data USA 2022 and FL Business Analytics Programs Data, 2022). The data was reliable based on the objective of the research.

3.1 Analysis of commuter colleges in Florida

From a sample of 10 community colleges in Florida, the researchers evaluated the level of total enrollment for a period of three years from 2017 to 2019 looking closely at the available Business Analytics courses available in the stated universities.

As shown in Table 1 (in Appendix), it is clear that Hillsborough Community College had the highest average enrollment at 24,403 students. None of the community colleges offered a Business Analytics course in their portfolio of available courses. Chipola, North Florida college, Gulf Coast State College, and Tallahassee Community College have no Business Analytics *(BA)* related courses at all. Despite the lack of BA-related courses in the aforementioned colleges above, distinct colleges such as Seminole State offer related courses like Statistics, Data Science, and Actuarial Science. Hillsborough Community College offers the related program Business Intelligence as well as a college credit certificate for BI. Whereas Eastern FL College offers courses in Database management Systems, Geographical Information Systems, and Data science.

Table 2 (in the Appendix) gives a summary of the available related courses and average college total enrollment. As shown in Table 2 we can note that most commuter colleges only offer related subjects with regard to Business Analytics, and as such in order to develop a skill base on BA one has to bridge to acquire more wholesome knowledge on the said field, hence there's evidence of a notable skill gap as student enrollment in the above-colleges will either need additional training to cover for the knowledge gap to step into the fast-growing industry of Business Analytics.

3.2 Analysis of large universities in Florida

A table was constructed via the collected enrollment data of the 10 sampled universities. As shown in Table 3 (in appendix) the 2017-2019 enrollment for each university was averaged to give a comparison to the average enrollment of the 10 sampled commuter colleges. The average enrollment for universities is slightly higher than that of commuter colleges as seen in Table 3 with a general average of 48,302 students in the period 2017 to 2019. The University with the highest average total enrollment is the University of Central Florida at 68,093 average total enrollment from 2017-2019. The University of South Florida and the University of Florida offer courses in Business Analytics and Information Systems, and Miami Dada College also offers a related course but in Business Intelligence. 2 out of the sampled 10 universities, Valencia College and Palm Beach State College do not offer any courses in relation to Business Analytics. Table 4 (in the Appendix) illustrates the diversity of courses for each sampled University. Of the 10 sampled universities, each one of them offers courses in either Business Analytics, Business Administration or Business Intelligence and Information Systems. Table 5 (In the Appendix) illustrates the count of each course overall in the sample of 10 universities. As shown in Table 5, despite the availability of Business Analytics-related courses at these universities, the majority only offer Masters Level courses. Table 6 (In the Appendix) illustrates that training at the bachelor's level follows closely with the certificates, with minor and associate's programs at the bottom of the list.

3.3 Assessment of overall data

From the assessment of the collected data, it was evident across both the big institutions and the commuter institutions that determining the domain under which business analytics falls was a challenge. This was mainly attributed to the fact that the subject areas where business analytics is taught varied across the different institutions. Some institutions offered it in statistics and operations research while others offered it under management information systems. A major challenge with this is the fact that the various subject areas apply business analytics within their discipline but do not necessarily teach the base knowledge needed to perform business analytics. To avert this, it would be necessary that all institutions ensure that any of their business analytics programs or majors should include an application course within each discipline using business analytics and that all business analytics majors should practice their craft within a secondary business subject as a minor, a second major or simply their interest in a subject area. Additionally, it would make it possible for each discipline to offer their majors and minors the opportunity to experience the application of business analytics within their field.

3.4 Finding 1 and Hypothesis 1

To assess hypothesis 1 " *If a curriculum developed for commuter colleges could close the knowledge and skill gap in terms of how Business Analytics is taught among commuter colleges and universities"*, a frequency distribution of courses organized by type of university and course category was created. This analysis will also assist in the research and development of Business Analytics programs for commuter colleges.

The frequency distribution as shown in Table 7 (in Appendix) was constructed in which each row corresponds to a course. The courses classified into categories were taken from the web portals of the 20 sampled Florida Universities & Colleges. The courses were classified into the following categories: Business Analytics, Business Intelligence, Databases, Visualization, and Others. Figure 1 (in the Appendix) is a bar chart visualization of Table 7's frequency distribution. Together these descriptive statistics display that the most frequent course categories in Large Universities are Databases, Business Analytics, and Business Intelligence. The most frequent course category for Commuter Colleges is Database courses. The Figure 2 pie chart (in the Appendix) is a distribution analysis of the sample data showing that most of the course types are taught in Large Universities/Colleges and only 11% are taught in commuter colleges.

Overall the recommended courses for commuter colleges should buffer against the missing and less frequent course categories. The missing and less frequent courses at commuter colleges are Business Analytics, Business Intelligence, and Visualization. We recommended, in order to cover this basis in terms of learning, that commuter schools build a Business Analytics curriculum that covers these bases. By covering the less frequent learning bases, It is concluded that a curriculum developed by commuter colleges could close the knowledge and skill gap between commuter colleges and universities in terms of the delivery of Business Analytics courses. Thus, the research question is answered in the affirmative.

3.5 Finding 2 and Hypothesis 2

Regarding hypothesis 2 "*Do diverse programs like Business Analytics have the possibility of increasing Student Retention at Commuter Colleges? "* a growth analysis was done to compare large university and commuter college sample growth. Figure 3 and Figure 4 (In Appendix) were constructed in which each bar is split corresponding to multiple enrollment years for the 20 sampled Florida learning institutions. The bar charts evaluated the level of total enrollment for a period of three years from 2017-2019. In contrast to the enrollment growth in this period, large institutions (Figure 3) have positive growth overall. Commuter institutions (Figure 4) have negative growth overall. Having more choices is a big factor in student retention since students have all their learning needs satisfied in one place.

Comparing Figure 3 to Figure 4, the Large University sample (denoted by Figure 3) was more than likely to have diverse program options such as Business analytics, while the small/commuter schools (denoted by figure 4) were less than likely to have the choice of a Business analytics programs. Thus we can infer that offering more diverse programs like Business Analytics have the possibility of increasing Student Retention at Commuter Colleges. As a result, the research question "*Do diverse programs like Business Analytics have the potential to increase student retention at Commuter Colleges*" is answered in the affirmative.

4. Discussion

This study sought to assess the impact and effectiveness of bringing business analytics courses to small Florida colleges by benchmarking the approaches used by different colleges in Florida that have created programs in business analytics. And to further determine the impact and effectiveness, a common benchmark was used to compare programs and we sought to understand if the programs fall within a common framework.

There is a notable difference between big Institutions and commuter institutions with regard to the availability of Business-related courses offered. Big Institutions tend to have the inclusion of the latter much more likely than their commuter counterparts. However, there is still limited data to make an informed conclusion concerning the extent that commuter institutions need Business Analytics. Also, there is a need to expand the sample size and review other projects to also assess possible factors that would influence learning institutions to venture into such business-related courses or find a way of augmenting the already available related course to take into account Business management and hence equip those who enroll that suit the market needs to better solve and propel businesses in the economy.

From the findings of the study, it is evident that there is a need for all colleges to include business intelligence and analytics in their curricula in a uniform way as the disjointed manner that was evident from the analysis will result in inconsistent preparation due to varying core knowledge. From these findings, we expect that smaller colleges should either develop a niche program with depth or develop a more generalized program with an inclination toward business analytics. Lastly, considering the cross-functional nature of business analytics, there is a need for cross-functional interaction between faculty in the design and implementation of courses. To ensure that the programs are able to build depth, there is a need that multiple stakeholders design and implement the programs.

5. Conclusion

This study sought to assess the impact and effectiveness of bringing business analytics courses to small Florida colleges. From an analysis of the data and literature on the subject, we were able to understand the current state of business analytics programs and how they are being developed. While the findings of this study were inconclusive about what business analytics curriculum should incorporate, there was a general consensus regarding the most important elements that each program should have to impart the relevant specific skills that are relevant to the job market. The college's board of trustees, which is responsible for program implementation, could essentially use the recommended fundamental course types as a checklist to ensure the program covers all of the bases, or if the institution already has a BA program, they could determine what is missing to improve it. Institutions' boards of trustees, who decide program implementation decisions, could simply glance at the researchers' 3-Year-Enrollment Growth graphs in Figures 5 and 6 (in Appendix) and compare them to their growth and program enrollment numbers to evaluate whether diversity affects their retention. According to evaluation results, schools might prioritize diversity in their meeting preparations. There are a variety of options available to educators. However, there is no doubt that developing a business analytics program requires a wide and deep understanding of the field if informed decisions are to be made about the content.

References

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., ... & Widom, J. (2011). Challenges and opportunities with Big Data 2011-1. Cyber Center Technical Reports. Paper 1. http://docs.lib.purdue.edu/cctech/1

Albright, C. S., & Winston, L. W. (2016). Business Analytics: Data Analysis & Decision Making. Cengage Learning. https://books.google.com/books?id=pSEADAAAQBAJ&dq=business+analytic+program&source=gbsnavlinkss

Carroll, D., Ryan, K., & Elliott, S. (2021, April 21). Want to improve your chances of getting a full-time job? A double degree can do that. https://theconversation.com/want-to-improve-your-chances-of-getting-a-full-time-job-a-double-degree-can-do-that-157306

Community College Review. (2021). Best community colleges in Florida (2021-22). Community College Review - Profiles of USA Community Colleges. https://www.communitycollegereview.com/florida

Data USA. (2022). The Enrollment Data for the Florida learning institutions. https://datausa.io/search/?q=&dimension=University

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different. MIT Sloan Management Review

Delen, D. Tatoglu, E., Bayraktar, E., Zaim, S., & Aydiner, A. S., (2019). Business analytics and firm performance: The mediating role of business process performance. Journal of business research, 96, 228-237.

Fitzgerald, M. (2015). Training the Next Generation of Business Analytics Professionals. MIT Sloan Management Review, 56(2), MIT Sloan management review, 2015-01-01, Vol.56 (2).

FIU. (n.d.). Mba in Business Analytics curriculum. MBA in Business Analytics Curriculum FIU Business. Retrieved December 1, 2021, from https://business.fiu.edu/graduate/business-analytics-mba/

Flbog. (2022). State University System of Florida Universities. Retrieved April 28, 2022, from https://www.flbog.edu/universities/

Fldoe. (2021). FLORIDA COLLEGE SYSTEM FTE ENROLLMENT. Florida Department of Education. Retrieved January 9, 2022, from https://www.fldoe.org/accountability/data-sys/CCTCMIS/reports.stml

Fldoe. (2022). College directory. Florida Department of Education. Retrieved April 28, 2022, from https://www.fldoe.org/accountability/data-sys/CCTCMIS/report-coordinator/college-directory.stml

Forrester. (2017, July 10). Hadoop is data's darling for a reason. https://www.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/

Freedman, M. (2020). How Businesses are Collecting Data (And What They're Doing with It). Business News Daily. https://www.businessnewsdaily.com/10625-businesses-collecting-data.html

FSU. (n.d.). Business analytics, Information Systems & Supply Chain. Business Analytics, Information Systems & Supply Chain. Retrieved December 1, 2021, from https://business.fsu.edu/departments

Gartner (n.d.). Definition of Business Analytics - Gartner Information Technology Glossary. Gartner. Retrieved January 9, 2022, from https://www.gartner.com/en/information-technology/glossary/business-analytics

Gavin, M. (2019). Business Analytics: What it is and Why it's Important. Harvard Business School. https://online.hbs.edu/blog/post/importance-of-business-analytics?tempview=logoconvert

Ghani, N. A., Hamid, S., Hashem, I. A. T., & Ahmed, E. (2019). Social media big data analytics: A survey. Computers in Human Behavior, 101, 417-428.

Gooden, Joachim. (2021). FL Business Analytics Program Data. Florida Agricultural and Mechanical University. https://docs.google.com/spreadsheets/d/1TVUtxoTKRBkNr1iLr4V_cO0uS94x14Z1ylk7OqcyFjQ/edit?usp=sharing

Leonardi, P., & Contractor, N. (2018). Better people analytics. Harvard Business Review, 96(6), 70-81.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Moreira, J., Carvalho, A. C. P. L. F., & Horváth, T. (2019). A general introduction to data analytics. Hoboken, NJ: John Wiley & Sons, Inc.

Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. Information Systems and e-Business Management, 16(3), 479-491.

Shabbir, M. Q., & Gardezi, S. B. W. (2020). Application of big data analytics and organizational performance: the mediating role of knowledge management practices. Journal of Big Data, 7(1), 1-17. https://doi.org/10.1186/s40537-020-00317-6

Surdak, C. W., & Agarwal, S. (2014). The Benevolent Side of Big Data: Data analytics can be used to drive growth in the developing world. Finance & Development, 51(004).

Appendix

Note: The data from the charts/figures are sourced from Florida institution web portals. Sources are linked in the excel document linked below:

Gooden, Joachim. (2021). *FL Business Analytics Program Data*. Florida Agricultural and Mechanical University. https://docs.google.com/spreadsheets/d/1TVUtxoTKRBkNr1iLr4V_cO0uS94x14Z1ylk7OqcyFjQ/edit?usp=sharing

Table 1: Commuter College Average Total Enrollment

AVERAGE COMMUTER COLLEGE TOTAL ENROLLMENT FOR THE PERIOD 2017 TO 2019					
UNIVESITY	**2017**	**2018**	**2019**	**AVERAGE ENROLLMENT**	**STANDARD DEVIATION (SD)**
Chipola College	2,092	2,081	1,939	2,037	69.68
Gulf Coast State College	5,379	5,644	4,797	5,273	353.77
North Florida College	1,324	1,367	1,229	1,307	57.66
Tallahassee Community College	24,639	18,458	18,401	20,499	2,927.28
Northwest Florida State College	5,727	2,497	5,025	4,416	1,387.10
Daytona State College	13,970	13,737	13,430	13,712	221.14
Eastern Fl State College	15,769	15,820	15,352	15,647	209.63
Indian River State College	17,598	16,686	16,942	17,075	384.07
Seminole State College	17,549	18,104	17,754	17,802	229.14
Hillsborough Community College	27,625	23,182	22,403	24,403	2,300.15

Table 2: Available Related Courses and Average Enrollment

COLLEGE	DEGREES OFFERED	AVERAGE ENROLLMENT
Chipola College	None	2,037
Gulf Coast State College	None	5,273
North Florida College	None	1,307
Tallahassee Community College	None	20,499
Northwest Florida State College	Computer Information Specialist	4,416

Daytona State College	Database Technology AS	13,712
Eastern Fl State College	Data Science Specialization, Oracle Certified Database Administration, Geographic Information Systems (GIS), and Computer Information Technology	15,647
Indian River State College	Computer Information Data Specialist	17,075
Seminole State College	Business and Information Management, Statistics, Data Science and Actuarial Science	17,802
Hillsborough Community College	information technology and Business Intelligence Specialist, Statistics	24,403

Table 3: University Average Total Enrollment

AVERAGE UNIVERSITY TOTAL ENROLLMENT FOR THE PERIOD 2017 TO 2019					
UNIVERSITY	**2017**	**2018**	**2019**	**AVERAGE ENROLLMENT**	**STANDARD DEVIATION (SD)**
University of South Florida Main Campus	43540	43846	44246	43,877	289.07
Florida State University	41362	41269	42450	41,694	536.15
Broward College	40754	40784	38976	40,171	845.32
Florida Atlantic University	30208	29772	30061	30,014	181.12
Florida International University	56718	57942	58711	57,790	820.68
Miami Dade College	56001	54973	51679	54,218	1,843.51
University of Central Florida	66183	68571	69525	68,093	1,405.61
University of Florida	52669	52218	52407	52,431	184.92
Valencia College	44834	46521	47940	46,432	1,269.59
Palm Beach State College	30,052	31,816	31,289	31,052	739.34

Table 4: Available Programs for each University

UNIVERSITY	LINE OF COURSE
University of South Florida Main Campus	Business Analytics & Information Systems
	Business Analytics
Florida State University	Business Analytics
	Business Administration
Broward College	Business Analytics
Florida Atlantic University	Business Analytics
Florida International University	Business Analytics
Miami Dade College	Business Intelligence
University of Central Florida	Business Analytics
University of Florida	Business Analytics & Information Systems
Valencia College	None
Palm Beach State College	None

Table 5: Count of Line of Course

Line of Course	Count of Line of Course
Business Administration	1
Business Analytics	11
Business Analytics & Information Systems	8
Business Intelligence	2

Table 6: Count of Program

LEVEL	Count
Master	7
Bachelor of Science	6
Certificate	4
Minor	3
Associate	2
Other	2

Table 7: Course Frequency in the 20 Sampled Universities/Colleges

Type of University	Course category	Frequency
Large Institutions	Business Analytics	28
	Business Intelligence	13
	Databases	49
	Other	3
	Visualization	3
Commuter Colleges	Business Intelligence	3
	Databases	9

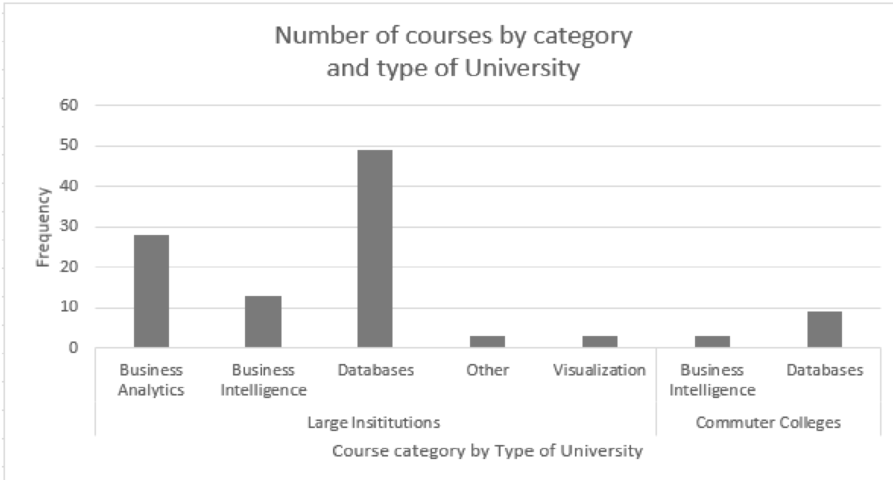


Figure 1: Descriptive Visualization of Course Frequency in the 20 Sampled Universities/Colleges

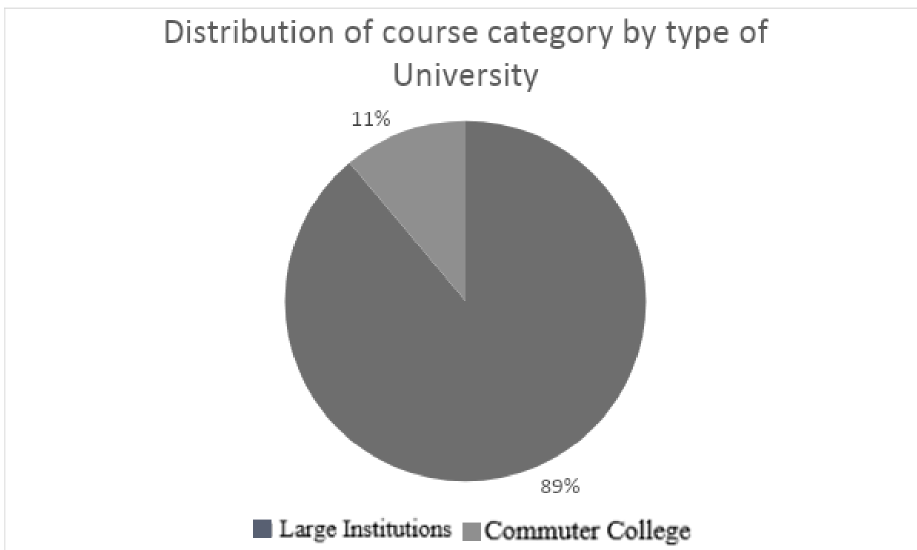


Figure 2: Distribution of course category by type of Institution

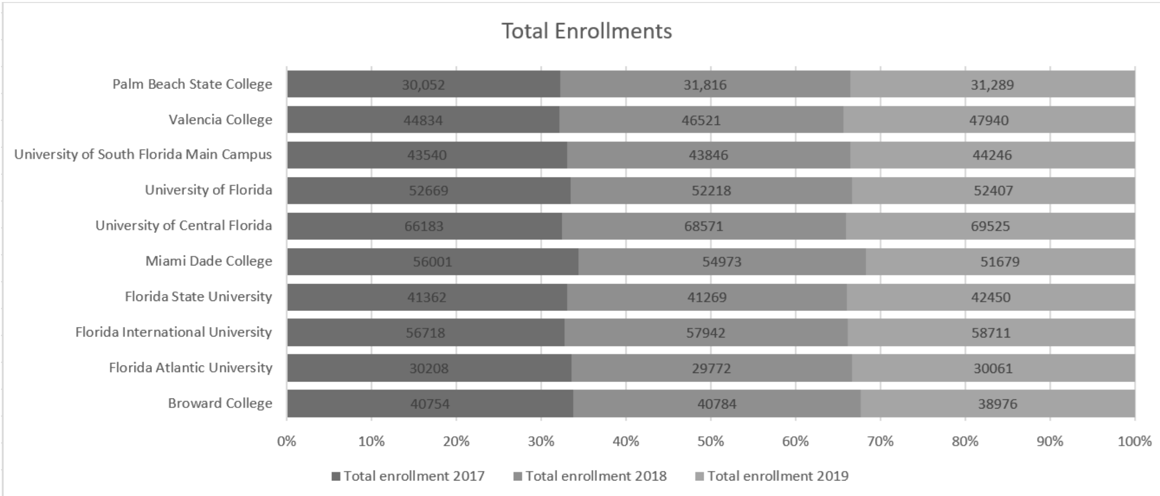


Figure 3: 3-Year Enrollment Growth of Large University sample

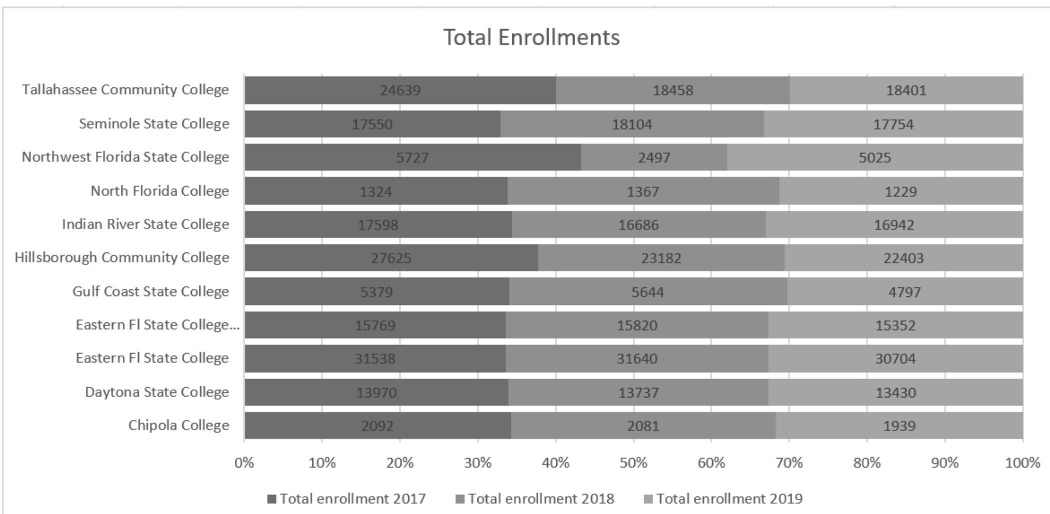


Figure 4: 3-Year Enrollment Growth of Commuter College sample

Effects of Functional Nanomaterials for Reduction of Carbon Monoxide Levels in the Workplace and Homes

Eylem Asmatulu*
Anh K. Pham
Department of Mechanical Engineering
Wichita State University, USA.
e.asmatulu@wichita.edu

Abstract

Carbon monoxide (CO) at high concentrations is extremely poisonous to humans and other invertebrates. It is one of the leading causes of unintentional deaths in the United States. Furthermore, people with heart diseases are more vulnerable to high CO levels since they already have a reduced ability to get oxygenated blood to their hearts. Functional nanomaterials (metallic nanoparticles [NPs], carbon nanotubes [CNTs], and nanofibers [NFs]) offer a high surface area, low weight-to-volume ratio, flexibility, and high porosity, which are perfect properties for gas absorption. Many nanomaterials are currently playing an essential role in CO-capturing technologies. This paper focuses on the usage of those functional nanomaterials for CO absorption and their absorption mechanisms. Due to high surface area and porosity, these materials have been widely used to absorb pollutants, including toxic gas like CO. Fabrications, CO-absorbing mechanisms, and absorbing properties of some nanomaterials will be discussed here.

Keywords: Nanomaterials, CO capture, and removal, absorbing mechanisms

1. Introduction

Carbon monoxide (CO) is a by-product of incomplete combustion. At high concentrations, CO can cause hypoxia (lack of oxygen at the tissue level) in humans and other vertebrates because once CO enters the bloodstream through the lungs, it immediately binds to hemoglobin with an affinity of 250 times more than oxygen, thereby preventing less oxygen being transported to tissue in the body. Depending on the concentration of CO, the victim can experience headaches, dizziness, nausea, unconsciousness, respiratory arrest, brain damage, or even death (Goldstein, 2008). Furthermore, CO at certain levels is vulnerable to people with heart disease since a limited amount of oxygenated blood goes to their hearts. CO is colorless, odorless, and nonirritating, and therefore particularly dangerous because the victim cannot sense it without a particular gas detector. According to the Centers for Disease Control and Prevention (CDC) (CDC, 2019), every year in the United States, at least 430 deaths and 50,000 emergency visits account for CO poisoning. Most poisoning events take place during vehicle incidents and house fires; however, in some cases, due to faulty ventilation, smoke containing CO from wood- or coal-burning stoves, generators, gas engines, natural gas equipment, or heating systems can be trapped in place, build up over time, and can cause a threat.

Therefore, it is important to have effective ways of capturing and removing CO. Of the several methods for removing CO, catalytic oxidation is the most effective. During the past few years, in addition to conventional materials and methods used to remove CO and improve CO reduction, the implementation of nanomaterials for CO removal has been reported. Nanomaterials are materials that are smaller than 1 µm and have at least one dimension less than 100 nm. In their study of nanomaterials, researchers have concluded that, compared with bulk materials, nanomaterials can possess unusual and

Submitted May 13, 2022
Revised August 28, 2022

extraordinary properties for the following reasons: (i) characteristic length scale, (ii) surface, and interface effect, and (iii) quantum size effect. Nanomaterials and nanotechnology have emerged in many fields of scientific, industrial, medical, and commercial sectors, with hundreds of thousands of products, such as filtration membranes, coatings, batteries, and so on (Abedin et al., 2021; R. Asmatulu & Khan, 2018; Hughes & Asmatulu, 2021; Sengul & Asmatulu, 2020). With the innovation of nanotechnology in the last two decades, many promising nanomaterials can be fabricated and utilized to treat toxic CO successfully. Among the many methods used to remove CO, catalytic oxidation in the presence of a catalyst is the most effective.

This review paper provides a brief introduction to some promising nanomaterials that have been used in recent years for CO catalytic oxidation. Materials with great surface area and high porosity, including metallic nanoparticles (NPs), carbon nanotubes (CNTs), electrospun nanofibers (NFs), or a matrix comprised of two or more of these, are great candidates for this application due to their absorption and high CO catalytic oxidation. A summary of all these nanomaterials including their size, function, and catalytic properties is listed in Table 1.

Table 1. Summary of Nanomaterials Used for CO Catalytic Oxidation

Nanomaterial	Size	Type of Catalyst	Catalytic Properties	Ref.
Pt NPs	Particle diameter: 2.2–16.7 nm	Pt/TiO_2	• Pt/TiO_2 (R)-600 had highest dispersion (33.4%) • CO conversion of Pt/TiO_2 (R)-600 was 100% at 25°C, the highest conversion rate compared to other samples	(Abedin et al., 2021)
	Cluster: 0.75 ± 0.11 nm	$Pt\text{-}Cu/CeO_2$	• Pt nanocluster deposited on Cu-O-Ce interface with activated oxygen lattice achieved 100% CO conversion rate at 30°C	(Liu et al., 2020)
Pd NPs	Particles: 2.7–5 nm	Pd/triptycene-based	• Pd loading to microporous remarkably enhanced the catalytic activity of CO oxidation • 3% loading of Pd exhibited the highest catalysis since it provided the most active sites; a 100% conversion rate of 3% Pd was at 160°C	(Liang et al., 2013)
	Particles: ~10 nm	Pd/Graphene	• Large Pd NPs created nanoholes on graphene sheets; that defect structure facilitated CO • 100% CO conversion rate was at 190°C	(Kumar et al., 2015)
	Bimetallic NPs: 3–5 nm	Pd-M/AC	• The addition of metallic NPs improved the catalysis of CO oxidation by providing more surface area • Pd/Cu-AC had the highest conversion rate of 50% at 30°C	(Singhania & Gupta, 2018)
	Particles: 3.5– 13.7 nm	Pd/TiO_2 nanosheet	• TiO_2 nanosheet provided a large surface area for Pd NPs, then created strong	(Zhai et al., 2018)

Nanomaterial	Size	Type of Catalyst	Catalytic Properties	Ref.
			electronic interaction TiO_2-Pd, which promoted catalysis. • 100% CO conversion rate at 80°C for Pd/TiO_2-350 sample.	
	Particles: 2.5–10.6 nm	Pd/Co_3O_4	• Pd NPs with the smallest size 2.5 nm yielded the highest dispersion 39.4%, which led to the highest catalytic activity • The formation energy of CO_2, endothermic, peaked at 85°C to 157°C from smallest to largest Pd NPs, indicating that smaller Pd NPs correlated to higher CO oxidation	(Huang et al., 2019)
Au NPs	Cluster: 1–6 nm dia	Au/TiO_2	• High catalytic activity on surface Au/ TiO_2 at 300 K (27°C) and 10 Torr • TOF peaked at 0.25 s^{-1} per site for 2–3 nm Au cluster diameter	(Valden et al., 1998)
	Hemispherical NPs: 3.8 ± 0.6 nm	Au/ TiO_2	• Highly sensitivity of CO on the perimeter interface of Au NPs • Calcined Au- TiO_2 had a conversion rate of CO to CO_2 of 80%	(Du et al., 2014)
	Particles: 2 nm	Au/POM	• Smaller Au particles (< 2 nm) resulted in higher catalytic activity. • The catalyst's function is affected by trace water • 100% CO conversion rate occurred within -40 °C to 60°C	(Yoshida et al., 2018)
CNTs	–	Ag/CNT	• The improved interface between Ag sites and CNTs since more Ag particles deposited on CNTs surfaces; prevented the sintering of Ag • 100% conversion rate at 170°C	(Wang, 2016)
	–	Cu-Mn/MWCNTs	• Cu-Mn synergistic interaction and properties of MWCNTs resulted in remarkable catalytic activity • 100% conversion rate at 180°C	(Guo et al., 2016)
	–	CuO-CeO_2/ MWCNTs	• High surface area and great electrical properties of MWCNTs enhanced synergistic interaction between CuO-CeO_2 and support • Alkalis improved the dispersion of Cu and Ce oxides, which facilitated catalysis • 100% conversion rate at 175°C	(Guo et al., 2016)
	–	CuO-CeO_2/MWCNTs	• Excellent catalytic activity due to more surface lattice oxygen and oxygen	(Shi et al., 2018)

Nanomaterial	Size	Type of Catalyst	Catalytic Properties	Ref.
			vacancies, which resulted from the high surface area provided by MWCNTs • 100% conversion rate at 120°C	
NFs	Fiber diameter: 9.8 nm	PVA/chitosan	• CO adsorption on PVA/chitosan nanofibers could be related to hydrogen bond formation and dipolar interaction • CO adsorbed was 14 mmol/g at 1.4 bar pressure and 26.85°C	(Sargazi et al., 2019)
	Fiber diameter: 200 nm	Au/CeO_2 NFs	• High Ce content dispersed on NFs suitable to fabricate excellent CeO_2 support for Au catalyst • The high aspect ratio of NFs induced randomly packed bed with higher porosity • CO conversion rate was 85% at 75°C	(Calderón et al., 2015)
	Fiber diameter: 0.5–0.67 nm	Co/K-OMS-2 NFs	• Redox capacity of the catalyst correlated with a high surface area of K-OMS-2 NFs • CO conversion rate was 100% at 100°C	(Yang et al.,2017)
	Fiber diameter: 92, 108, 133 nm	RuO_x/CeO_2 NFs	• The high surface area of NFs enhanced catalytic activity • Adsorption sites for CO and O_2 increased; correlated with a hollow interior and porous exterior of CeO_2 NFs • ~ 90% conversion rate below 150°C	(Liu et al., 2020)

2. Nanomaterials for CO Sorption

2.1 Noble metal nanoparticles

The oxidation of CO to CO_2, which is a less poisonous gas, has always drawn great attention in many scientific fields because it is the most effective method for removing CO that is used in air purification, three-way catalytic conversion in automobiles, and pollution control. For the past few decades, noble metals such as gold (Au), platinum (Pt), palladium (Pd), and ruthenium (Ru) have been widely studied and practically used in the CO catalytic oxidation process due to their remarkable properties because they resist both chemical and thermal reactions, and they can accelerate oxidation. These noble nanomaterials are tuned explicitly for converting CO to CO2. The gas adsorption and separation can be increased by generating more active metal centers (e.g., exposed metal cations) that can functionalize the ligands by incorporating functional groups in the organic ligands to increase or tune the guest-host interaction.

Recently, nanotechnology has been studied extensively, resulting in a renewed interest in noble metal catalysts at the nanoscale. These metal NPs are promoted to enhance the oxidation process at lower (room) temperatures because of their fascinating properties as nanostructures. Noble metal NPs can increase the beneficial metal-support interactions because they can maximize the population of active centers, resulting in better cost-effectiveness and robust catalytic materials (Wang et al., 2020).

Figure 1 shows the mechanism of CO oxidation in the catalytic system. The fundamental concept of this mechanism involves metal oxide support, which provides oxygen for the CO oxidation reaction, and noble metal NPs act as a catalyst to facilitate the binding of CO on the catalyst surface to facilitate oxidation. Noble metal NPs are Pt, Pd, Au, etc. Metal oxides, which are rich in oxygen, are commonly used in Fe_2O_3, CeO_2, Mn_3O_4, Co_3O_4, and so on.

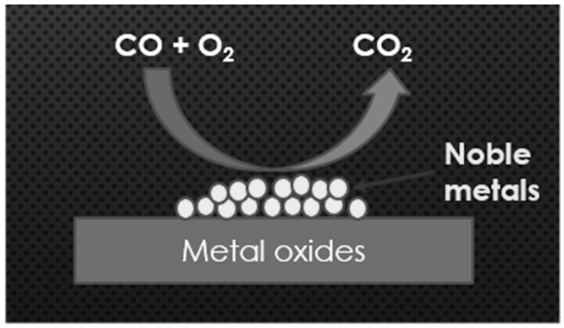


Figure 1. CO catalytic oxidation on a noble catalyst.

2.1.1 Platinum nanoparticles

Platinum is the most common noble metal used in the catalytic process in both research and commercial application. Pt has many d-orbital electrons, which facilitate its binding to other molecules so that the oxidizing reaction can be facilitated. Related to the quantum effect, the size of Pt particles has become a research interest of many scientists due to their role in catalysis in the CO oxidation process. For instance, Kim and Jhi (2011) studied the CO oxidation catalytic performance of Pt NPs on graphene support and observed that with defects on the surface of graphene, more CO was tolerated on Pt NPs than on Pt bulk material. Many recent studies have shown that modified Pt catalysts can catalyze CO oxidation at low temperatures, especially when accompanied by reducible oxides like TiO_2, CeO_2, or SiO_2. For example, Kim et al. (2016) used TiO_2 as support for a mixture of the Pt metallic (PtO) and Pt valence (PtO_x) states (Pt^{2+}, Pt^{4+}) with different NP sizes, and reported that a higher Pt dispersion was associated with smaller Pt particles at room temperature when Pt/TiO_2 catalysts were prepared and heat-treated with H_2 at different temperatures. They further discovered that at room temperature, only Pt^0 sites responded to the activation of CO better than Pt^{2+} sites; hence, higher CO oxidation activity was followed by a higher value of Pt metallic/Pt total. Liu et al. (2020) reported a strategy to activate low-temperature reactions using modified Pt-Cu-CeO_2 catalysts. They controlled the size of Pt and used oxide doping to precisely tune the interfacial reducibility and structure of the catalyst. The Cu-O-Ce interface with activated lattice oxygen accompanied by Pt subnanoclusters showed the ability of the Pt-Cu-CeO_2 catalyst to facilitate the CO oxidation at low temperatures. Furthermore, Li and his co-workers (2021) reported that not all Pt particles were active during the CO and H_2O catalytic process, only those at the perimeter of the interface of Pt NPs and SiO_2 support. Therefore, understanding NPs and their positions on the catalyst may help remove the inactive ones in order to minimize NP usage, thereby improving cost-effectiveness. This catalytic mechanism can also be implemented in many other catalytic systems (Li et al., 2021).

2.1.2 Palladium nanoparticles

As one of the metals in the platinum groups, palladium also possesses great properties for CO oxidation catalysis. In fact, Pd along with rhodium are vital components of automobile catalytic converters. Pd NPs have also gained the attention of many researchers relative to CO oxidation catalysis. According to Liang et al. (2013), when Pd NPs were well dispersed on a triptycene-based microporous polymer, the catalytic system showed excellent CO oxidizing since the porous structure provides stability for the Pd NPs. Many studies have demonstrated the influence of NP size of Pd on the CO catalytic process (Burange et al., 2018; Christou & Efstathiou, 2007; Haneda et al., 2017; Z. Wang et al., 2010). For example, Huang et al. (2019) prepared Pd/Co_3O_4 in different sizes (2.5–10.6 nm) and observed that when different sizes of Pd NPs were dispersed on the Co_3O_4 support, smaller Pd NPs correlated with the increase in CO oxidation catalytic activity. However, particle size is not the only factor that affects the CO oxidation process. Zhai et al. (2018), in their study of the CO oxidation on Pd with TiO_2 nanosheet support, showed that higher catalytic activity and better thermal stability of Pd/TiO_2 resulted from the electronic interaction between TiO_2 spheres and Pd NPs. Pd NPs could also form nanoholes on the graphene sheet, which improved the CO oxidation as well as hydrogen storage because it increased the rate of Pd NP interdiffusion (Kumar et al., 2015). Singhania and Gupta (2018) reported that the dispersion of bimetallic Pd-M particles (M = Cu, Ni, and Co) with an average size of around 3–5 nm could contribute to excellent catalytic activity at low temperatures for CO oxidation; among all of them, the Pd/Cu-AC catalyst had the best conversion rate of 50% at 30^oC. However, one of the challenges of Pd NPs used in catalysis is the rapid agglomeration of those particles; thus, to improve the stability of Pd NPs in catalysis, more studies need to be conducted, which requires considerable effort, time, and money (Zhai et al., 2018).

2.1.3 Gold nanoparticles

Being seen as far less catalytic compared to platinum and platinum metal groups, gold nanoparticles appear to have a remarkable effect on the CO oxidation reaction catalysis. This phenomenon was studied by Haruta (1997) and Haruta et al. (1987) who reported that Au NPs, around 5 nm in size, were able to oxidize CO below 0^oC. This fascinating property of Au NPs then captured tremendous attention from catalytic science. For instance, Valden et al. (1998) prepared Au NPs of different sizes from 1 to 6 nm and investigated the relationship between the thickness of Au particles deposited on TiO_2 and the catalysis of CO oxidation. The highest turnover frequency (TOF) of catalytic activity was approximately 0.25 s^{-1} per site for bilayers of Au NPs with diameters of 2–3 nm. Du et al. (2014) also studied the influence of Au NPs sizes on CO oxidation using an Au-TiO_2 catalyst. They prepared a set of 2.9, 3.8, and 5.1 nm Au NPs to deposit on TiO_2 support, calcined Au-TiO_2, and observed that different sizes of Au particles resulted in other boundaries between Au NPs with TiO_2 support. This boundary was referred to as the interface perimeter, and the greater the interface perimeter, the more catalytic activity took place. According to the analysis, the optimum size yielding the greatest interface perimeter was 3.8 ± 0.6 nm, where the conversion rate reached 80% at 60^oC reaction temperature. Furthermore, Yoshida et al. (2018) mounted Au NPs with a diameter of around 2 nm on a polyoxometalate (POM) support to fabricate polyoxometalate-supported gold nanoparticulate catalysts. They observed stable catalysis of Au/POM with a 100% CO conversion rate over a wide range of temperatures (-40 ^oC to 60^oC). The modified catalyst also showed no degradation over 35 days at 0^oC.

2.2 Carbon nanotubes

Carbon nanotubes are interesting nanomaterials. Because of their nanosize (diameter varies from 0.4 to 40 nm), CNTs possess unusual mechanical strength and thermal and electrical conductivity; however, their properties of low thermal expansion coefficient, as well as high porosity and surface area, allow CNTs to have the ability to attach molecules to their walls. In fact, CNTs have already been studied and practically used in some catalytic processes since the large surface area of CNT can lead to higher deposition of metal NPs and enhance the catalytic cycle. For instance, Wang (2016) dispersed Ag NPs on the surface of CNTs, resulting in a higher interface between the Ag particles and CNTs. This protected Ag NPs from agglomeration, thus providing more adsorption sites for CO oxidation. Guo et al. (2016) impregnated copper (Cu) and manganese (Mn) on multi-wall carbon nanotubes (MWCNTs) to facilitate the CO oxidation catalytic process by enhancing the synergistic interaction of bimetallic Cu/Mn and properties of MWCNTs. MWCNTs can be employed as support for the $CuO-CeO_2$ catalyst for preferential CO oxidation. According to Gao et al. (2015), the interaction between CuO and CeO_2 was strengthened due to their high dispersion on MWCNTs, and the catalytic activity of $CuO-CeO_2$/MWCNTs catalyst became pronounced compared to other supporting materials. Shi et al. (2018) also reported that CuO and CeO_2 dispersed uniformly on the surface of MWCNTs, which enhanced the catalytic performance by providing more lattice oxygen and oxygen vacancies on the surface. Nevertheless, CNT usage in catalytic science is still a challenge because when combined with other molecules, a covalent bond (strong bond) is formed, and that can change the structure of the CNT. The main issue left for researchers to resolve for further improvement of CO oxidation using CNTs as support in the catalyst is developing a method that respects the initial structure of CNTs and forms a covalent bond between the CNTs and other gas molecules.

2.3 Nanofibers

Electrospun nanofibers are versatile materials, and have numerous practical applications, including filtration, tissue engineering, drug delivery, biosensor, fuel cell, and solar cell, due to their extraordinary properties, which are high surface area, high porosity, low weight-to-volume ratio, especially flexibility, and, in some NF-modification cases, great mechanical strength (Uddin et al., 2021). Some of the properties of electrospun NFs can be utilized to serve a role in the mechanism of CO reduction. Sargazi et al. (2019) developed polyvinyl alcohol (PVA)/chitosan electrospun nanofibrous membrane to adsorb CO, and according to their testing model, the capacity of adsorption of this material was reported to be higher than activated carbon, zeolite, and even a metal-organic framework. Another interesting application is that electrospun nanofibers can be used as a support for the catalyst to facilitate CO oxidation since nanofibers can provide a large surface area and have high porosity.

Calderón et al. (2015) reportedly used carbon nanofibers (CNFs) as support for Pt-Ru. They observed that introducing Ru to Pt-based CNFs enhances the CO oxidation activity at 20–70^{o}C, since there are more particles deposited on the surface of CNFs, and some oxygenated groups from CNFs are responsible for promoting CO oxidation. Eid et al. (2019) also implemented one-dimensional carbon nitrile-based nanofibers (1D gC3N–4) as support for the Au-Pd catalyst. As described, 1D gC3N–4 had unique physicochemical properties that could be synergic with the outstanding catalytic merits of Au-Pd thus creating a superior catalytic mechanism for CO oxidation. Yang et al. (2017) developed a system called cobalt-doped K-OMS-2 nanofibers in which cobalt was doped with manganese octahedral molecular sieves, which changed the K-OMS-2 structure from nanorods to nanofibers; then the framework was reported to complete the CO oxidation process at 100^{o}C in the presence of moisture. Z. Liu et al. (2020) reported fabricating CeO2 NFs by electrospinning, then loading RuO_x on the modified NFs to create RuOx-CeO_2 NF catalysts. The new catalyst had good thermal stability and excellent CO conversion at low temperatures (~90% below 150^{o}C) due to the high surface area provided by the NFs. With their extraordinary properties, NFs have become great implementations for catalysis, in the past decades, as catalytic support with uniform dispersion and stability on a large surface.

Moreover, the cost aspect and scale-up of the catalyst with NF support might become less complicated when using the electrospinning method because of the ability to disperse NPs uniformly. Nonetheless, there are still drawbacks to using NFs as support since their high-temperature resistance is not advanced compared to other metallic or CNT support (E. Asmatulu et al., 2022). More attention and studies are needed to utilize these nanomaterials in the CO catalytic reduction process successfully.

The disposal of nanomaterials generally follows the same pathways as conventional waste. Before ending up in a landfill or incineration, nanoparticles can be thermally/chemically/physically/biochemically processed to reduce possible adverse effects on human health and the environment. Or discharged nanoparticles can be recycled through some processes (magnetic, flotation, electrostatic, chemical leaching, etc.).

Nanotechnology is a growing and emerging field. However, because it is new, protocols for safe handling and disposal are incomplete and will need more research funding, and studies to develop proper protocols for nanoparticle recycling (Khan et al., 2022).

3. Conclusion

In brief, some nanomaterials such as Au, Pt, Pd, or their combination in the form of nanoparticles, carbon nanotubes, nanofibers, or the matrices of these possess great properties to enhance the CO catalytic oxidation reaction, which is part of the CO-reduction mission in environmental and health improvement. Nanomaterials have several advantages over bulk materials because when materials are scaled down to nanometers, they own unique properties. Not all catalyst sites are active; therefore, it is necessary to minimize the materials and maximize catalytic activation, and nanomaterials can potentially solve this problem. In addition, NPs, CNTs, and NFs have a relatively higher surface area than bulk materials with the same volume, which can promote better CO adsorption and oxidation catalysis. The research exploration of nanomaterials for CO reduction, compared to that of other materials, is still in the burgeoning stage; therefore, researchers see nanomaterials as great opportunities for further studies in this field of application. Nonetheless, nanomaterials and technologies for CO catalytic oxidation have many challenges in terms of cost, stability of materials, and oxidation rate of CO on those materials. This will require the effort and dedication of researchers as well as investment in this field.

References

Abedin, F., Asmatulu, E., & Andalib, M. N. (2021). Nanomaterials and human health: An overview. In *Nanotoxicology and Nanoecotoxicology Vol. 2*, 165–180. Springer.

Asmatulu, E., Andalib, M. N., Subeshan, B., & Abedin, F. (2022). Impact of nanomaterials on human health: A review. *Environmental Chemistry Letters*, 1–21.

Asmatulu, R., & Khan, W. S. (2018). *Synthesis and Applications of Electrospun Nanofibers*. Elsevier.

Burange, A. S., Reddy, K. P., Gopinath, C. S., Shukla, R., & Tyagi, A. K. (2018). Role of palladium crystallite size on CO oxidation over CeZrO4-δ supported Pd catalysts. *Molecular Catalysis*, *455*, 1–5.

Calderón, J. C., García, G., Querejeta, A., Alcaide, F., Calvillo, L., Lázaro, M. J., Rodríguez, J. L., & Pastor, E. (2015). Carbon monoxide and methanol oxidations on carbon nanofibers supported Pt–Ru electrodes at different temperatures. *Electrochimica Acta*, *186*, 359–368.

Centers for Disease Control and Prevention. (2019, January 25). Picture of America: Poisoning. https://www.cdc.gov/pictureofamerica/pdfs/Picture_of_America_Poisoning.pdf

Christou, S. Y., & Efstathiou, A. M. (2007). Effects of Pd particle size on the rates of oxygen back-spillover and CO oxidation under dynamic oxygen storage and release measurements over Pd/CeO2 catalysts. *Topics in Catalysis*, *42*(1), 351–355.

Du, M., Sun, D., Yang, H., Huang, J., Jing, X., Odoom-Wubah, T., Wang, H., Jia, L., & Li, Q. (2014). Influence of Au particle size on Au/TiO2 catalysts for CO oxidation. *The Journal of Physical Chemistry C*, *118*(33), 19150–19157.

Eid, K., Sliem, M. H., Eldesoky, A. S., Al-Kandari, H., & Abdullah, A. M. (2019). Rational synthesis of one-dimensional carbon nitride-based nanofibers atomically doped with Au/Pd for efficient carbon monoxide oxidation. *International Journal of Hydrogen Energy*, *44*(33), 17943–17953.

Gao, Y., Xie, K., Wang, W., Mi, S., Liu, N., Pan, G., & Huang, W. (2015). Structural features and catalytic performance in CO preferential oxidation of CuO–CeO 2 supported on multi-walled carbon nanotubes. *Catalysis Science & Technology*, *5*(3), 1568–1579.

Goldstein, M. (2008). Carbon monoxide poisoning. *Journal of Emergency Nursing*, *34*(6), 538–542.

Guo, Y., Lin, J., Li, C., Lu, S., & Zhao, C. (2016). Copper manganese oxides supported on multi-walled carbon nanotubes as an efficient catalyst for low temperature CO oxidation. *Catalysis Letters*, *146*(11), 2364–2375.

Haneda, M., Todo, M., Nakamura, Y., & Hattori, M. (2017). Effect of Pd dispersion on the catalytic activity of Pd/Al2O3 for C3H6 and CO oxidation. *Catalysis Today*, *281*, 447–453.

Haruta, M. (1997). Size- and support-dependency in the catalysis of gold. *Catalysis Today*, *36*(1), 153–166. https://doi.org/10.1016/S0920-5861(96)00208-8

Haruta, M., Kobayashi, T., Sano, H., & Yamada, N. (1987). Novel gold catalysts for the oxidation of carbon monoxide at a temperature far below 0°C. *Chemistry Letters*, *16*(2), 405–408. https://doi.org/10.1246/cl.1987.405

Huang, R., Kim, K., Kim, H. J., Jang, M. G., & Han, J. W. (2019). Size-controlled Pd nanoparticles loaded on Co3O4 nanoparticles by calcination for enhanced CO oxidation. *ACS Applied Nano Materials*, *3*(1), 486–495.

Hughes, S., & Asmatulu, E. (2021). Nanotoxicity and nanoecotoxicity: Introduction, principles, and concepts. In *Nanotoxicology and Nanoecotoxicology Vol. 1*, 1–19. Springer.

Khan, W.S., Asmatulu, E., Uddin, M.N. and Asmatulu, R. (2022). Recycling and Reusing of Engineering Materials: Recycling for Sustainable Developments. Paperback ISBN: 9780128224618

Kim, G. J., Kwon, D. W., & Hong, S. C. (2016). Effect of Pt particle size and valence state on the performance of Pt/TiO2 catalysts for CO oxidation at room temperature. *The Journal of Physical Chemistry C*, *120*(32), 17996–18004.

Kim, G., & Jhi, S.-H. (2011). Carbon monoxide-tolerant platinum nanoparticle catalysts on defect-engineered graphene. *ACS Nano*, *5*(2), 805–810.

Kumar, R., Oh, J.-H., Kim, H.-J., Jung, J.-H., Jung, C.-H., Hong, W. G., Kim, H.-J., Park, J.-Y., & Oh, I.-K. (2015). Nanohole-structured and palladium-embedded 3D porous graphene for ultrahigh hydrogen storage and CO oxidation multifunctionalities. *ACS Nano*, *9*(7), 7343–7351.

Li, Y., Kottwitz, M., Vincent, J. L., Enright, M. J., Liu, Z., Zhang, L., Huang, J., Senanayake, S. D., Yang, W.-C. D., & Crozier, P. A. (2021). Dynamic structure of active sites in ceria-supported Pt catalysts for the water gas shift reaction. *Nature Communications*, *12*(1), 1–9.

Liang, Q., Liu, J., Wei, Y., Zhao, Z., & MacLachlan, M. J. (2013). Palladium nanoparticles supported on a triptycene-based microporous polymer: Highly active catalysts for CO oxidation. *Chemical Communications*, *49*(79), 8928–8930.

Liu, X., Jia, S., Yang, M., Tang, Y., Wen, Y., Chu, S., Wang, J., Shan, B., & Chen, R. (2020). Activation of subnanometric Pt on Cu-modified CeO2 via redox-coupled atomic layer deposition for CO oxidation. *Nature Communications*, *11*(1), 4240. https://doi.org/10.1038/s41467-020-18076-6

Liu, Z., Lu, Y., Confer, M. P., Cui, H., Li, J., Li, Y., Wang, Y., Street, S. C., Wujcik, E. K., & Wang, R. (2020). Thermally Stable RuO x–CeO2 Nanofiber Catalysts for Low-Temperature CO Oxidation. *ACS Applied Nano Materials*, *3*(8), 8403–8413.

Sargazi, G., Afzali, D., Mostafavi, A., Shadman, A., Rezaee, B., Zarrintaj, P., Saeb, M. R., Ramakrishna, S., & Mozafari, M. (2019). Chitosan/polyvinyl alcohol nanofibrous membranes: Towards green super-adsorbents for toxic gases. *Heliyon*, *5*(4), e01527.

Sengul, A. B., & Asmatulu, E. (2020). Toxicity of metal and metal oxide nanoparticles: A review. *Environmental Chemistry Letters*, *18*(5), 1659–1683.

Shi, L., Zhang, G., & Wang, Y. (2018). Tailoring catalytic performance of carbon nanotubes confined CuOCeO2 catalysts for CO preferential oxidation. *International Journal of Hydrogen Energy*, *43*(39), 18211–18219.

Singhania, A., & Gupta, S. M. (2018). Low-temperature CO oxidation: Effect of the second metal on activated carbon supported Pd catalysts. *Catalysis Letters*, *148*(3), 946–952.

Uddin, M. N., Desai, F. J., Subeshan, B., Rahman, M. M., & Asmatulu, E. (2021). Sustainable atmospheric fog water generator through superhydrophobic electrospun nanocomposite fibers of recycled expanded polystyrene foams. *Surfaces and Interfaces*, *25*, 101169.

Valden, M., Lai, X., & Goodman, D. W. (1998). Onset of catalytic activity of gold clusters on titania with the appearance of nonmetallic properties. *Science*, *281*(5383), 1647–1650.

Wang, F. (2016). Effect of support carbon materials on Ag catalysts used for CO oxidation in the presence and absence of H2. *Journal of Environmental Chemical Engineering*, *4*(4), 4258–4262.

Wang, Z., Li, B., Chen, M., Weng, W., & Wan, H. (2010). Size and support effects for CO oxidation on supported Pd catalysts. *Science China Chemistry*, *53*(9), 2047–2056.

Wang, G., Lu, K., Yin, C., Meng, F., Zhang, Q., Yan, X., Bing, L., Wang, F., & Han, D. (2020). One-Step Fabrication of PtSn/γ-Al2O3 Catalysts with La Post-Modification for Propane Dehydrogenation. *Catalysts*, *10*(9), 1042.

Yang, J., Zhou, H., Wang, L., Zhang, Y., Chen, C., Hu, H., Li, G., Zhang, Y., Ma, Y., & Zhang, J. (2017). Cobalt-Doped K-OMS-2 Nanofibers: A Novel and Efficient Water-Tolerant Catalyst for the Oxidation of Carbon Monoxide. *ChemCatChem*, *9*(7), 1163–1167.

Yoshida, T., Murayama, T., Sakaguchi, N., Okumura, M., Ishida, T., & Haruta, M. (2018). Carbon Monoxide Oxidation by Polyoxometalate-Supported Gold Nanoparticulate Catalysts: Activity, Stability, and Temperature- Dependent Activation Properties. *Angewandte Chemie*, *130*(6), 1539–1543. https://doi.org/10.1002/ange.201710424

Zhai, X., Liu, C., Chang, Q., Zhao, C., Tan, R., Peng, H., Liu, D., Zhang, P., & Gui, J. (2018). TiO 2-nanosheet-assembled microspheres as Pd-catalyst support for highly-stable low-temperature CO oxidation. *New Journal of Chemistry*, *42*(22), 18066–18076.

Designing, Fabricating, and Testing of Portable Electrospun Nanofilter Systems for Domestic Water Purification

Kiran Nepal[1]
Waseem S. Khan[2*]
Altaf Khan[2]
Ramzan Asmatulu[1*]
[1]*Department of Mechanical Engineering, Wichita State University, USA*
[2]*Department of Mechanical Engineering, Fujairah Men's College, UAE*
wkhan1@hct.ac.ae; ramazan.asmatulu@wichita.edu

Abstract

Membrane technology has been playing a vital role in water filtration and purification. Various technologies such as sand filtration, reverse osmosis, treatment with chemical disinfectants, and distillation have been used in the past for clean water sources. Membrane technology is a relatively easy method with many important features, including low operational temperature, low power consumption, and free of chemical additives. Membrane filtration can be further improved by using nanofibrous media in portable forms for drinking water supplies. Electrospun nanofibers possess high porosities, high surface area, good permeability and flexibility, unidirectional orientation, and well-connected pore structures, and therefore, an ideal candidate for portable water filtration purposes. Electrospinning is a novel method of fabricating nanofibers in various forms. In this study, nanofibrous membranes were fabricated using two polymers (polyacrylonitrile - PAN and polyethylene glycol -PEG), one solvent (dimethylformamide - DMF), and activated charcoal at different electrospinning conditions (DC voltages, pump speeds, and collector distances). Different water samples from the local pond, water jet cutter water, tap water, deionized (DI) water, and carbon black (CB) deionized water were used in this study. The water samples were tested for turbidity, total dissolved solids (TDS), electrical conductivity (EC), pH, and refractive index. The test results indicated that the proposed portable nanofilter system substantially improved the quality of the water sources for drinking and other household use. It is concluded that these nanofibrous membranes can easily purify contaminated water by eliminating impurities, such as organic and inorganic particles, viruses, bacteria, algae, and other pathogens.

Keywords: Wastewater, Electrospinning, Nanofibrous Membrane, Water Treatment, Drinking Water.

1. Introduction

Nanomaterials are relatively new with many potential applications in various industries (medicine, aerospace, energy, textile, defense, electronics, and environment), and have been used in wastewater treatment and drinking water production for a while. Polymeric nanofiber membranes have been receiving enormous attention due to their applications in filtration technology, biotechnology, and many other new fields (Wendorff et al., 2012; Colvin, 2003; Dolez et al., 2009). The idiosyncratic advantage of membrane technology is its high efficiency, low power consumption, and minimum investment. Electrospun nanofibrous membranes possess a high surface area to volume ratio, high porosity, tunable pore size, and ease of fabrication, which make them ideal candidates in filtration and membrane science. Electrospinning is one of the simplest and most uncomplicated techniques for generating nanosized fibers from polymeric solutions (Asmatulu et al., 2013; Nepal, 2018; Hammel et al., 2004).

Submitted June 20, 2022
Revised August 28, 2022

Water is an essential element in human life. Lack of pure water in developing countries creates many health problems, such as diarrhea, cholera, typhoid, dysentery, malaria, jaundice, and waterborne disease; therefore, it is very important that available water must be pure and domestically used. Nanotechnology presents the potential for the treatment of surface water, groundwater, wastewater, and air by toxic metal ions, organic and inorganic substances, and microorganisms. The research and development of polymeric nanofibers have gained much interest in recent years due to their wide range of applications in engineering and medical science. Compared to conventional microfibers, polymeric nanofibers possess high filtration efficiency due to higher inertia impaction and interception. Nanofibers are usually produced by the electrospinning process, a valuable technique based on the use of electrostatic forces to produce continuous nanofibers with a diameter of about 10 nanometers to some micrometers.

Nanofiber can remove impurities such as suspended particles, viruses, bacteria, mold, fungus, algae, and other pathogens from water. Electrospun nanofibers possess high surface area, high porosity, good permeability, nanosized diameter, and well-connected pore structures, which are suitable for filtration technologies (Wang et al., 2007). Although polymeric nanofibers have been used in air filtration, their applications in water treatment are yet to be explored. However, laboratory experiments with electrospun nanofibers for water treatment have shown some promising results, but their commercial applications have not been fully investigated. The membrane fabricated using an electrospinning process are generally referred to as electrospun nanofibrous membranes (Ghorani, 2015; Yoshikawa & Isezaki, 2014; Alarifi et al., 2016; Alarifi et al., 2015; Asmatulu & Khan, 2019; Nepal, 2018).

The main objective of this research was to design a portable nanofiller system using an electrospinning process for removing bacteria and other harmful impurities from water and purifying it for drinking purposes. The first step in this research was to produce the electrospun nanofibrous membrane using two polymers such as polyacrylonitrile (PAN), polyethylene glycol (PEG), and one solvent (dimethylformamide – DMF). The goal here was to evaluate different layers of filtration and reduce the total dissolved solids, pH, electrical conductivity, and turbidity and increase the refractive index of water. After filtration, each parameter was compared to evaluate the performance of different layers of nanofibers.

2. Experiment

Two polymers (PAN and PEG), one solvent (DMF), carbon black nanoparticles (~50 nm), and activated charcoal particles (below 1 mm in diameter) were purchased from Sigma Aldrich and used without any alterations or purification. The electrospinning process was used to fabricate the nanofibrous membranes. Electrospinning is a versatile method to process solutions or melts, mainly of polymers into continuous fibers with diameters ranging from a few micrometers to a few nanometers. First, the polymer was dissolved in a solvent at 60°C and 700 rpm, and then the solution was transferred to a 10 ml syringe mounted on a syringe pump. In this study, 9.5 wt. % of polyacrylonitrile, 0.5 wt.% of polyethylene glycol, and 90 wt.% of dimethylformamide were used to make a polymeric solution. An electrode was inserted into the plastic syringe and the other end of the electrode was connected to a high-voltage DC supply. The applied voltages were maintained at 25 kV and the distance between the tip of the plastic syringe and the collector screen was maintained at 25 cm. The pump speed was kept constant at 1 ml/hour throughout the tests. These test parameters were determined based on the trial-and-error method. The collector screen was grounded. Finally, the nanofiber was kept in an oven for 30 min by applying heat at 90°C. Heat treatment on the nanofiber helps to improve its mechanical properties and chemical stability. A Thelco Laboratory oven was used for heat treatment. Figure 1 shows SEM images of the same nanofiber without and with heat treatment, respectively.

From the SEM images shown in Figure 1, the morphology of the nanofiber changes with the heat treatment processes, which may be associated with the realignment of the polymeric chains in the nanofibers. After the heat treatment, the thermal and chemical stability of fibers could also be improved (Mohammad et al., 2017). No beads and pores formation were observed in SEM images, which is particularly good for water filtration. We assume that the thermal treatment is necessary since fibers should be chemically stable and mechanically strong for the portable filtration application.

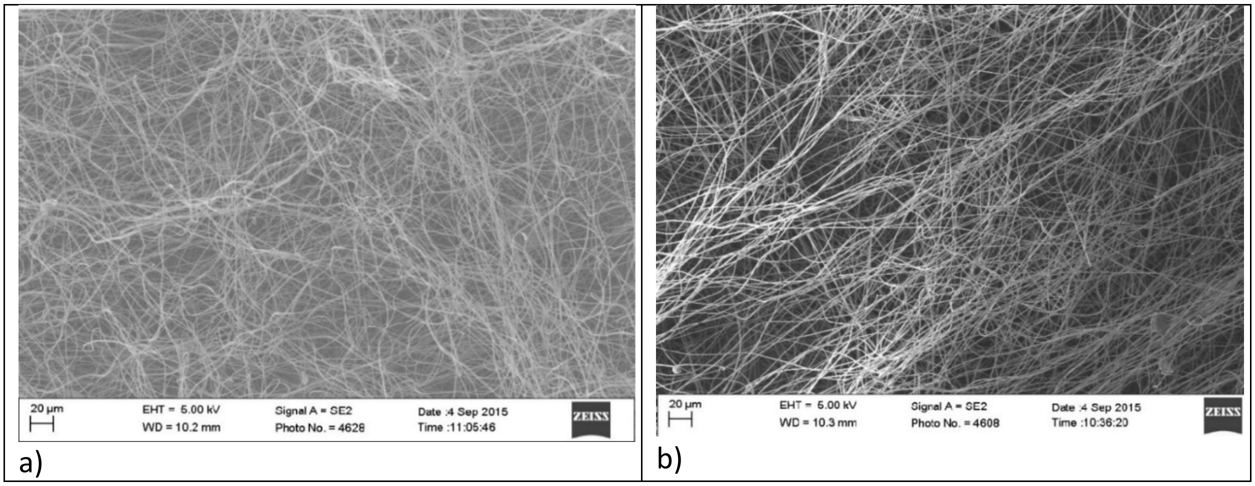


Figure 1: the SEM images of the same nanofiber membranes a) without and b) with heat treatment processes.

3. Results and Discussion

The different water samples were filtered using three different layers of nanofiltration systems. The total dissolved solids, electrical conductivity, turbidity, pH, and refractive index of different water samples are discussed for each sample.

3.1 Treating Pond Water

Pond water is one of the most widely available water resources, however, it is not exempt from impurities that can hinder its use in domestic and industrial applications. These impurities include dust, fine sand, clay, dirt, metal pieces, and biological contaminants. Pure water is tasteless, colorless, and odorless, but pond water sources, rainfall, streams, and wells, are known to contain harsh chemicals and elements, including acid, ions, and metallic rust elements. Among all sources, pond water contains many physical and chemical contaminants based on the locations (industrial and agricultural zones, oil, and chemical facilities) (Asmatulu et al., 2013).

Table 1 shows the different water parameters for the pond water. After one, two, and three layers of filtration systems combined with the activated charcoal, three layers nanofiber filtration system shows much better test results compared to other filtration options. The pH of the collected pond water was found to be 8.70, which is slightly high for drinking purposes, but after filtration, the pH was reduced to 8.57 for one layer of nanofiber, 8.36 for two layers of nanofiber, and 8.2 for three layers of nanofiber. The total dissolved solids of the pond water were 610 ppm before filtration, but after filtration, the TDS was decreased to 248 ppm for one layer of nanofilter, 225 ppm for two layers of nanofiber, and 219 ppm for three layers of nanofiber, which are all acceptable for drinking purposes. The electrical conductivity of the pond water was 1220 µs/cm prior to the filtration, and after the filtration, the electrical conductivity values were decreased to 496 µs/cm, 450 µs/cm and 438 µs/cm for one layer, two-layer and three-layer of nanofiber membranes, respectively. The turbidity of pond water was 18 NTU before water filtration, and after the filtration, the turbidity was decreased to 4.8 NTU, 4.4 NTU, and 3.8 NTU for one-layer, two-layer, and three-layer membranes, correspondingly. This indicates that the produced water is acceptable for drinking purposes. The refractive index studies of the pond water were also conducted on the same samples. It was 1.3329 before the tests, and slightly increased to 1.3331 and 1.3332 after the filtration which may be related to the dispersion of particles and clusters, but it is still acceptable to drink based on the standards [4].

Table 1: Different water quality parameters for the pond water

Parameter	TDS (ppm)	EC (µs/cm)	Turbidity (NTU)	Ph	Refractive Index
Before Filtration	610	1220	18	8.7	1.3329
One Layer Nanofiber	248	496	4.8	8.57	1.3331
Two Layers Nanofiber	225	450	4.4	8.36	1.3331
Three Layers Nanofiber	219	438	3.8	8.2	1.3332

3.2 Treating water jet cutter water

About 500 ml of contaminated water was collected from a water jet cutter at National Institute for Aviation Research, Wichita, Kansas, USA. This water sample contained many impurities, which can easily be seen by the naked eye. The pH of water from the water jet cutter was found to be 8.6, which was high for drinking purposes, but after filtration, the pH was reduced to 8.45 for one layer of nanofiber, 8.31 for two layers of nanofiber, and 8.15 for three layers of nanofiber. The total dissolved solids of water from the water jet cutter were around 435 ppm before the filtration and after the filtration, it was decreased to 246 ppm, 232 ppm, and 220 ppm for one-layer, two-layer, and three-layer nanofilter membranes, correspondingly. The electrical conductivity of the same water was initially 870 µs/cm and decreased to 492 µs/cm, 464 µs/cm, and 440 µs/cm after the one, two, and three layers of nanofiber membranes, which are acceptable for drinking purposes at this condition. The turbidity tests were also conducted, and its value was reduced from 64 NTU to 4.7 NTU after three layers of nanofibers. The refractive index of water from the water jet cutter was 1.3327 before filtration, it was very similar after the layers of the filtration process.

Based on the test results, it is concluded that these treated water sources are acceptable for drinking purposes. Table 2 shows the different parameters of water from the water jet cutter before and after the filtration. As can be seen in Table 2 that three layers filtration system presents much better performance, and all parameters are within the acceptable range for drinking water. Also, no significant changes were observed in the water quality after a few weeks of shelf life (storage time in the lab conditions).

Table 2: Different water quality parameters for water jet cutter water

Parameter	TDS (ppm)	EC (µs/cm)	Turbidity (NTU)	pH	Refractive Index
Before Filtration	435	870	64	8.6	1.3327
One Layer Nanofiber	246	492	4.7	8.45	1.3329
Two Layers Nanofiber	232	464	4.4	8.31	1.3331
Three Layers Nanofiber	220	440	4.1	8.15	1.3332

3.3 Treating deionized water

Table 3 shows the different parameters for deionized water. Deionized water has almost all its minerals removed; therefore, its total dissolved solids and electrical conductivity are near zero. The pH of the collected deionized water was found to be 6.09-6.68, which was lower for drinking purposes, but after filtration, the pH was increased to 8.46 for one layer of nanofiber, due to the presence of the activated charcoal. The pH was decreased to 8.30 for two layers of nanofiber, and 8.25 for three layers of nanofiber.The TDS was almost 0 ppm before filtration because of the lack of salt and other ions, but after filtration, the TDS was increased to 13 ppm for one layer of nanofilter, 16 ppm for two layers of nanofiber, and 18 ppm for three layers of nanofiber, which may not be acceptable for drinking purposes. The electrical conductivity of deionized water was 0 µs/cm before filtration, but after filtration, the electrical conductivity was increased to 26 µs/cm for one layer of nanofiber, 32 µs/cm for two layers of nanofiber, and 36 µs/cm for three layers of nanofiber, which are not acceptable for drinking purposes. The turbidity of the deionized water was 0 NTU before filtration, but after filtration, turbidity was changed to 1.4 NTU for one layer of nanofiber, 1.1 NTU for two layers of nanofiber, and 0.9 NTU for three layers of nanofiber. The refractive index of deionized water was 1.3329 before filtration but after filtration, no significant changes were observed in the refractive index.

Table 3: Different water quality parameters for the deionized water source

Parameter	TDS (ppm)	EC (µs/cm)	Turbidity (NTU)	pH	Refractive Index
Before Filtration	0	0	0	6.68	1.3327
One Layer Nanofiber	13	26	1.4	8.46	1.3328
Two Layers Nanofiber	16	32	1.1	8.30	1.3329
Three Layers Nanofiber	18	36	0.9	8.25	1.3331

3.4 Treating tap water

This water sample (table 4) was collected from a tap located in Wallace Hall Room 125 at WSU. The parameters of this tap water were good, but after nanofiltration, the water quality was even better. The pH of the tap water was found to be 8.2, but after filtration, the pH was reduced to 8.15 for one layer, 7.9 for two layers, and 7.7 for three layers of nanofiber systems. The total dissolved solids of the tap water were initially 367 ppm before the filtration, and after the three layers of filtration, it was reduced to 240 ppm for one layer, 225 ppm for two layers, and 205 ppm for three layers of nanofibers, which are acceptable for drinking purposes. The electrical conductivity of tap water was 734 µs/cm at the beginning and then reduced to 480 µs/cm for one layer, 450 µs/cm for two layers, and 410 µs/cm for three layers of nanofiber membranes. The turbidity of the tap water was 4.6 NTU and was decreased to 3.9 NTU for three layers of nanofiber, which are acceptable for drinking purposes. The refractive index did not change much after the filtration. Table 4 shows the different parameters for the tap water, whereas Figure10 shows the same tap water before and after the filtration.

3.6 Microfilter tests on different water samples

In this study, the performances of microfilter and nanofilter were compared, as well. Microfiltration removes suspended solids from water due to the small pore size of the filter. Unlike nanofiltration, microfilters remove unwanted larger particles ranging between 0.1 and 10 microns, including bacteria, but microfiltration cannot remove dissolved contaminants, unlike nanofiltration.

Table 4: Different water quality parameters for the tap water

Parameter	TDS (ppm)	EC (µs/cm)	Turbidity (NTU)	pH	Refractive Index
Before Filtration	367	734	4.6	8.2	1.3328
One Layer Nanofiber	240	480	4.4	8.15	1.3329
Two Layers Nanofiber	225	450	4.2	7.9	1.3331
Three Layers Nanofiber	205	410	3.9	7.7	1.3332

Recently, membrane filtration technologies, especially those carried out with microfiltration have shown promising results (Belibi et al., 2015; Li & Visvanathan, 2017) and 24 ppm for three layers of nanofibers. The electrical conductivity of carbon black deionized water was 10 µs/cm at the beginning but increased to 30 µs/cm for one layer, 36 µs/cm for two layers, and 48 µs/cm for three layers of nanomembrane systems. The turbidity of carbon black deionized water was 17 NTU before the water treatment and decreased to 4.1 NTU for three layers of nanofiber. The refractive index values of carbon black deionized water were 1.3324 before the filtration and were similar after the nanofiltration processes. Table 5 shows the different parameters for carbon black deionized water.

Table 5: Different water quality parameters for carbon black deionized water sources

Parameter	TDS (ppm)	EC (µs/cm)	Turbidity (NTU)	pH	Refractive Index
Before Filtration	5	10	17	6.76	1.3324
One Layer Nanofiber	15	30	4.7	8.5	1.3327
Two Layers Nanofiber	18	36	4.5	8.41	1.3328
Three Layers Nanofiber	24	48	4.1	8.35	1.3331

Most portable filters employ microfilters for water filtration. Since the micro filter has a pore size of 0.1–10 µm, which is larger than the nanofilter which has a pore size of 0.1 nm–0.001 µm (Manimekalai et al., 2017). In this research, the main aim was to show that water quality was improved using a nanofilter having a smaller pore size for a better quality of drinking water.

Table 6 shows different water parameters before and after microfiltration of different water samples. The different water samples were filtered using microfilters instead of nanofilters. When pond water was filtered using a microfilter, the TDS, EC, pH, and turbidity quality were not as good as in nanofiltration, especially with three layers of nanofilters. Before microfiltration, the TDS was 610 ppm, but after microfiltration, the TDS decreased to 390 ppm. This value was still high in comparison with the nanofiltration results. Before microfiltration, the electrical conductivity of pond water was 1220 µs/cm, but after microfiltration, the electrical conductivity was decreased to 780 µs/cm, which was still higher than the pond water filtered by the nanofilter. The electrical conductivity of pond water after three layers of nanofilters was 438 µs/cm. The turbidity and pH of pond water were 18 NTU and 8.7, respectively, before microfiltration, but after microfiltration, these values decreased to 5.2 NTU and 8.6, respectively. However, both values were still higher than pH and turbidity results after the nanofiltration systems. It can easily be concluded that the use of a nanofilter is much better than the use of a microfilter. In the case of water from the water jet cutter, the TDS, electrical conductivity, pH, and turbidity were 320 ppm, 640 µs/cm, 8.5, and 6.2 NTU, respectively, after the microfiltration. With the nanofiltration, the TDS, electrical conductivity, pH, and turbidity were 220 ppm, 440 µs/cm, 8.15, and 4.1 NTU. This clearly shows that nanofiltration is much better than microfiltration.

Table 6: Different water parameters for different water samples before and after microfiltration

Parameter	Pond Water		Water Jet Cutter Water		Tap Water	
	Before Microfiltration	After Microfiltration	Before Microfiltration	After Microfiltration	Before Microfiltration	After Microfiltration
TDS (ppm)	610	390	435	320	367	300
EC	1220	780	870	640	734	600
P^H	8.7	8.6	8.6	8.5	8.2	8.2
Turbidity (NTU)	18	5.2	64	6.2	4.6	4.5

Similarly, with the tap water, the TDS, electrical conductivity, pH, and turbidity were 300 ppm, 600 µs/cm, 8.2, and 4.5 NTU, respectively, after the microfiltration. However, these values were 205 ppm, 410 µs/cm, 7.7, and 3.9 NTU with the nanofiltration (three layers of nanofibers). Thus, the nanofiltration due to its small pore size is far better than the microfiltration systems. Reverse osmosis-having a pore size of less than 0.5 nm-is generally used in desalination and microfiltration-having a pore size between 100 and 10,000 nm-is used for bacterial removal. Ultrafiltration having a pore size between 2 to 100 nm is used in bacterial and virus removal (Muntha et al., 2017). Nanofiltration and reverse osmosis are pressure-driven processes. The application of nanofiltration lies between ultrafiltration and reverse osmosis; particularly, the nanomembrane retains organic molecules smaller than ultrafiltration membranes and has fewer monovalent ions compared to the reverse osmosis membrane (Ferrarini et al., 2001).

3.7 Fourier-Transform infrared spectroscopy analysis

The FTIR spectroscopy analysis was used to obtain an infrared spectrum of the absorptions of the three different polymer samples. Figure 2 shows the spectrums for PEG, PAN, and PAN/PEG nanofiber, respectively. As can be seen, the three samples show various peaks with different absorption characteristics. The functional group of nitriles is between wavenumbers 2100 and 2200. As FTIR spectroscopy was performed on the nanofiber, peaks started to form in the region between wavenumbers 2000 and 2500, which indicates the functional group of nitriles, meaning the presence of PAN in the nanofiber. The functional group of PEG is between wavenumbers 1000 and 1200. The peaks of PEG can also be seen in the nanofiber. A separate FTIR test was also done for PAN and PEG, and the peak of the individual wave number matched the FTIR peak done for the combined PAN/PEG polymer nanofiber. The FTIR spectra of PAN fibers have many peaks which are related to the existence of CH_2, C≡N, C=O, C−O, and C−H bonds (Farsani et al., 2009). A peak is observed at 2,916 cm^{-1}, which is related to C−H bonds in CH, CH_2, and CH_3 (Farsani et al., 2009; Alarifi et al., 2018, Alharbi et al., 2016).

3.8 Hydrophilic behavior of PAN nanomembrane

The hydrophilic and hydrophobic properties of polymer surfaces are determined by the water contact angle values between the water droplet and the solid polymeric fiber surface. A hydrophobic surface is one on which a droplet of water makes a contact angle greater than 90°, whereas a hydrophilic surface is

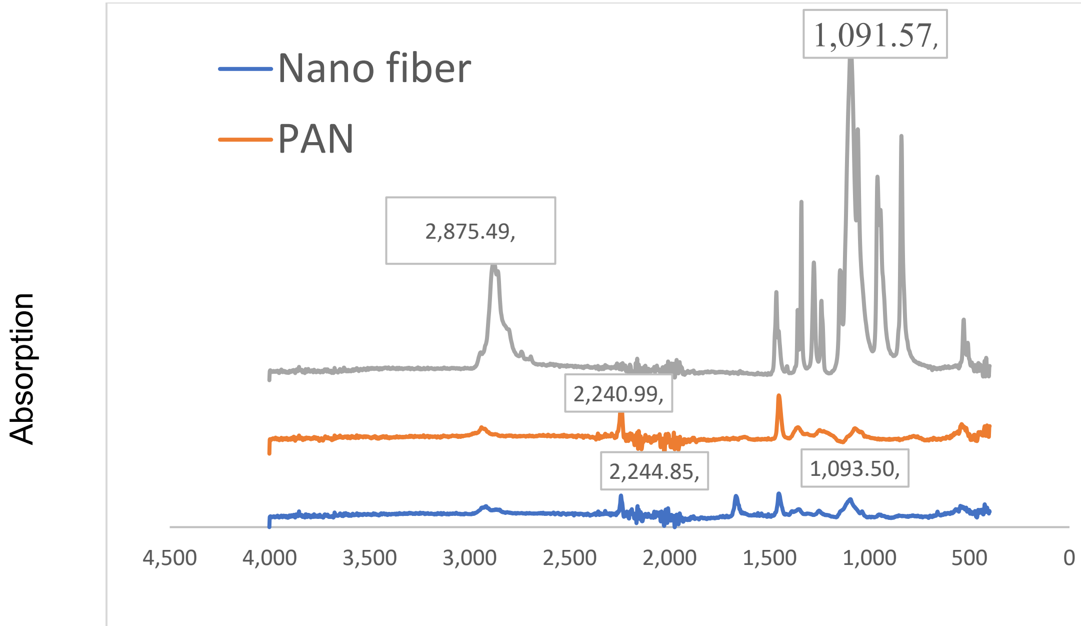


Figure 2: FTIR spectroscopy analysis of (a) PEG, (b) PAN, and (c) PAN/PEG nanofiber membranes.

one on which a droplet of water makes a contact angle less than 90°. Polymer surfaces with a contact angle between 150° and 180° are known as superhydrophobic. A hydrophilic surface has a strong affinity to water whereas a hydrophobic surface repels water A hydrophilic surface has a strong affinity to water whereas a hydrophobic surface repels water. Superhydrophobic surfaces are surfaces that exhibit a water contact angle of less than 10° (Uddin et al., 2018; Asmatulu, 2013). Our studies showed that PAN nanofibers form a contact angle of around 23-45° with water droplets. However, the addition of 0.5 – 5.0 wt.% of polyethylene glycol reduced the water contact angle below 10 °. The nanofiltration membrane used in this study has hydrophilic features and is therefore wet entirely during the filtration process. Hydrophilic surfaces are termed water-loving surfaces (Otitoju et al., 2017). Water lies on such surfaces as a flat film rather than rolling in the form of droplets. This ability of water plays an important role in many practical applications, including water treatment, water and oil separation, pervaporation, and biomedical applications (Mohammad et al., 2017; Jabbarnia et al., 2016; Asmatulu et al., 2011; Khan et al., 2013).

5. Conclusions

Portable nanofiltration is considered a promising separation technique to produce drinking water from different types of water sources. This filtration can reject several trace organic and inorganic compounds and heavy metals and ions with a lower energy demand than other processes for the production of high-quality drinking water in developed and under-developed countries. In this study, two polymers such as PAN and PEG were used to fabricate electrospun nanofiber membranes for treating water from various sources (tap water, pond water, deionized water, carbon black deionized water, and water from a water jet cutter). Three different types of filtration membrane systems were designed containing one, two, and three nanofiber layers of membranes. Activated carbon particles were also added to the nanofilter systems, which increased the surface area for the adoption of ions in the water. The water samples were tested for turbidity, total dissolved solids (TDS), electrical conductivity (EC), pH, and refractive index. FTIR spectra and water contact angle measurements were also carried out for the polymers and nanofibers used in this study. Test results indicated the three layers of hydrophilic nanofiber membranes provided the highest quality drinking water when compared to the other options.

References

Alarifi, I. M., Alharbi, A., Khan, W., Asmatulu, R. (2016). Carbonized electrospun polyacrylonitrile nanofibers as highly sensitive sensors in structural health monitoring of composite structures. *Journal of Applied Polymer Science*, 133(13), 43235-43241.

Alarifi, I. M., Alharbi, A.R., Khan, M.N., Khan, W.S., Usta, A., and Asmatulu, R. (2018). Water Treatment Using Electrospun PVC/PVP Nanofibers as Filter Medium. *International Journal of Material Science and Research*, 1(2), 43-49, 2018. DOI: 10.18689/ijmsr-1000107. https://madridge.org/journal-of-material-science-and-research/ijmsr-1000107.php

Alarifi. I. M. , Alharbi, A., Khan, W. S., Swindle, A., Asmatulu, R. (2015). Thermal, electrical and surface hydrophobic properties of electrospun polyacrylonitrile nanofibers for structural health monitoring. *Materials*, 8, 7017-7031.

Alharbi, A., Alarifi, I. M., Khan, W. S., and Asmatulu, R. (2016). Highly Hydrophilic Electrospun Polyacrylonitrile/Polyvinylpyrrolidone Nanofibers Incorporated with Gentamicin as Filter Medium for Dam Water and Wastewater Treatment. *Journal of Membrane and Separation Technology*, 5, 38-56.

Asmatulu, R. *Nanotechnology Safety*, Elsevier, Amsterdam, The Nederland, August 2013. https://www.elsevier.com/books/nanotechnology-safety/asmatulu/978-0-444-59438-9

Asmatulu, R., Ceylan, M., & Nuraje, N. (2011). Study of Superhydrophobic Electrospun Nanocomposite Fibers for Energy Systems. *Langmuir*, 27(2), 504-507. https://doi.org/10.1021/la103661c.

Asmatulu, R., Muppalla, H., Veisi, Z., Khan, W.S., Ceyla, M., Asaduzzaman, A., and Nuraje, N. (2013). Filtration of Micro and Nanosize Suspended Particles via Highly Hydrophilic Electrospun Nanofiber Membranes. The SAMPE Fall Technical Conference, Wichita, KS, USA, Oct. 21-24, 14.

Belibi, P. B., Nguemtchouin M.M.G., Rivallin M., Nsami, J. N., Sieliechi, J., Cerneaux, S., Ngassoum, M. B., Cretin, M. (2015). Microfiltration ceramic membranes from local Cameroonian clay applicable to water treatment. *Ceramics International*, 41 (2), Part B, 2752-2759. ISSN 0272-8842. https://doi.org/10.1016/j.ceramint.2014.10.090.

Colvin, V. L. (2003). The potential environmental impact of engineered nanomaterials. *Nature Biotechnology*, 21(10), 1166-1170.

Dolez, P. I., Bodila, N., Lara, J., and Truchon, G. (2009). Personal protective equipment against nanoparticles. *International Journal of Nanotechnology*, *7*(1), 99-117,

Farsani, R. E., Raissi, S., Shokuhfar, A., & Sedghi, A. (2009). FT-IR Study of Stabilized PAN Fibers for Fabrication of Carbon Fibers. *World Academy of Science, Engineering and Technology*, 50, 430-433. [PDF] FT-IR Study of Stabilized PAN Fibers for Fabrication of Carbon Fibers (researchgate.net)

Ferrarini, R., Versari, A., Galassi, S. (2001). A preliminary comparison between nanofiltration and reverse osmosis membranes for grape juice treatment. *Journal of Food Processing,* 50, 113-116.

Ghorani. B, Tucker, N., Yoshikawa, M. (2015). Approaches for the assembly of molecularly imprinted electrospun nanofiber membranes and consequent use in selected target recognition. *Food Research Journal,* 78, 448-464.

Hammel, E., Tang, X., Trampert, M., Schmitt, T., Mauthner, K., Eder, A., and Pötschke, P. (2004). Carbon nanofibers for composite applications. *Carbon*, 42(5), 2004, 1153-1158.

Jabbarnia, A., Khan, W. S., Ghazinezami, A., and Asmatulu, R. (2016). Investigating the Thermal, Mechanical and Electrochemical Properties of PVdF/PVP Nanofibrous Membranes for Supercapacitor Applications. *Journal of Applied Polymer Science,* 133, 43707. https://doi.org/10.1002/app.43707.

Asmatulu, R, Waseem S. Khan (2019). Synthesis and Applications of Electrospun Nanofibers. *Elsevier*. ISBN 978-0-12-813914-1.(https://www.elsevier.com/books/synthesis-and-applications-of-electrospun-nanofibers/asmatulu/978-0-12-813914-1)

Khan, W. S., Asmatulu, R., Ceylan, M., and Jabbarnia, A. (2013). Recent progress on conventional and non-conventional electrospinning processes. *Fibers and Polymers*, 14(8), 1235-1247. https://doi.org/10.1007/s12221-013-1235-8

Li, L. and Visvanathan, C. (2017). Membrane technology for surface water treatment: advancement from microfiltration to membrane bioreactor. *Reviews in Environmental Science and Bio/Technology*, 16, 737–760.

Manimekalai, D., Deepika, S., Brema, K., Rajeswari, C. and Juliet Selvarani, A. (2017). Microfiltration Technology in Wastewater Treatment. *International Journal of Current Research*, 9(03), 47310-47312. https://www.journalcra.com/article/microfiltration-technology-waste-water-treatment

Mohammad, S., Uddin, N., Hwang, G., and Asmatulu, R. (2017). Superhydrophobic PAN Nanofibers for Gas Diffusion Layers of Proton Exchange Membrane Fuel Cells for Cathodic Water Management. *International Journal of Hydrogen Energy*, 43, 11530-11538.

Muntha, S. T., Ayesha, K., Muhammad, S. (2017). Advances in Polymeric Nanofiltration Membrane: A Review. *Polymer-Plastics Technology and Engineering*, 56(8), 841-856.

Nepal, K. (2018). Design and Evaluation of Portable Nanofilter Systems for Continuous Water Supplies in Water-Scare Locations. M.S. Thesis, Wichita State University, Oct. 19.

Nepal, K. (2018). Design and Evaluation of Portable Nanofilter Systems for Continuous Water Supplies in Water-Scare Locations. M.S. Thesis, Wichita State University, Oct. 19.

Otitoju, T. A., Ahmad, A.L., and Ooi, B.S. (2017). Superhydrophilic (superwetting) surfaces: A review on fabrication and application. *Journal of Industrial and Engineering Chemistry,* 47, 19-40.

Uddin, M. N., Alamir, M., Muppalla, H., Rahman, M. M., and Asmatulu, R. (2018, December 34-24). Nanomembranes for Sustainable Fresh Water Production. The 5th International Conference on Mechanical Industrial and Energy Engineering, Khulna, Bangladesh. https://www2.kuet.ac.bd/icmiee2018/files/ICMIEE18-KN01_Keynote%20Paper_MMR-Uddin.pdf.

Wang, C., Yan, E., Huang, Z., Zhao, Q., Xin. Y. (2007). Fabrication of highly photoluminescent TiO2/PPV Hybrid nanoparticle-polymer fibers by electrospinning. *Macromolecular Rapid Communication*, 28, 205–209.

Wendorff, J.H., Agarwal, S., Greiner, A. (2012). *Electrospinning: Materials*, *Processing*, *and Applications*; Wiley: Singapore.

Jun Isezaki and Masakazu Yoshikawa. (2014). Molecularly imprinted nanofiber membranes: localization of molecular recognition sites on the surface of nanofiber. *Journal of Membrane and Separation* Technology, 3, 119-126. DOI:10.6000/1929-6037.2014.03.03.2

Investigating the Effects of Salt Spray Corrosion on Mechanical Properties of Steel Sheets of HVAC Heat Pumps Coated with Zinc and Polymeric Substances

Tyler P. McCune[1]
Waseem S. Khan[2*]
Ramazan Asmatulu[1 *]
[1]*Department of Mechanical Engineering, Wichita State University, USA*
[2]*Department of Mechanical Engineering, Fujairah Men's College, UAE*
wkhan1@hct.ac.ae; ramazan.asmatulu@wichita.edu

Abstract

Steel sheets are extensively used in various industrial applications because of their high mechanical properties, costs, and availability. However, in some applications such as gas, oil, and solar water heaters and heating, ventilation, and air conditioning (HVAC) systems, corrosion properties need to be further improved through new surface coatings for enhanced protection against corrosion, degradation, and abrasion. In many cases, a duplex coating system is used to provide extra protection against corrosion and degradation. In this study, steel sheet samples with varying levels of protective zinc coating thicknesses were individually tested as well as in conjunction with polymeric-based paint coatings through equal exposure times in a salt spray chamber for several days. Corrosion and mechanical testing of varying zinc coatings, especially in duplex systems are hardly available in the literature reviews, and this study reports new findings in the HVAC field. The prepared samples were kept in a salt spray chamber for different time frames (250 – 1000 hours), and the effects of corrosion damages over time were investigated. Before and after the evaluations, all the sheet samples were subjected to tensile testing to study the changes in mechanical properties. It was discovered that duplex coating systems that involve a zinc coating with an organic polymer-based powder coat or pre-painted finishes on metal sheets (G30, G60, and G90) showed little or no sign of corrosion degradation over time, and therefore experience substantially fewer property changes. However, the surfaces of the uncoated or one-layer coated samples were severely damaged in the salt spray chamber, reducing the lifetime of the HVAC system. It is concluded that the duplex system provided better corrosion and chemical resistance against environmental degradations.

Keywords: Corrosion Formation, Corrosion Prevention, Protective Coatings, Evaluation.

1. Introduction

Corrosion is a process of deteriorating materials or their surfaces due to environmental influences and is a very destructive attack (Asmatulu et al., 2007). Corrosion has a substantial environmental and economic impact on all facets of infrastructure, from bridges, highways, buildings, oil and gas pipes, chemical processing, water, and wastewater treatment systems. Besides material loss, corrosion interferes with the safety of humans, damages industrial operations, and poses a danger to the environment and health. Therefore, corrosion awareness and adaptation of timely and appropriate control measures are key to lessening corrosion failures during the service (Roberge, 2008).

Corrosion involves oxidation and reduction reactions on the surface of materials exposed to a harsh environment. These reactions take place at the anode and cathode sides. Oxidation occurs at the

Submitted June 21, 2022
Revised September 7, 2022

anode, reduction occurs at the cathode, and electrons follow from the anode to the cathode (Ahmad, 2006). This means that there is a loss of electrons due to the negative potential of the anode compared to the cathode. In short, the anode is the part of the metal that corrodes by changing its properties. The cathode, on the other hand, is considered nobler than the anode, and therefore gains reactions and does not corrode. This reaction takes place based on the potential difference between the anode and cathode and is carried out by an exchange of ions known as anions and cations (Ahmad, 2006). Anions are the negatively charged ions that move in a current toward the anode, while cations are the positively charged ions that do the opposite and move toward the cathode.

Conversion coatings are applied to galvanized steel surfaces to improve the adhesion of additional protective coatings and for the protection of the zinc coating (Hernandez-Alvarado et al., 2012). In this study, a duplex system of coatings has been used. A duplex system is a type of corrosion protection in which a powder coating or paint is applied to a steel surface that has undergone a hot-dip galvanizing process. When these two methods are used, the level of corrosion protection is much better than using each system independently. Powder coating or painting over hot-dip galvanized steel requires thorough preparation, proper skills, and clear knowledge of the two systems. The duplex system can be very useful when the powder coating or paint achieves good bonding, and there is correct substrate preparation of the galvanized surface (Hernandez-Alvarado et al., 2012).

Zinc coating is an economical and effective method for protecting steel substrates from the atmospheric corrosion environment. To facilitate cathode protection, zinc coating acts as a barrier against the corrosive environment by corroding itself sacrificially (Bhadu et al., 2013). Protection by galvanization may be enhanced by using a thicker zinc coating or by painting the metallic substrate (Guin, 2011). Paints generally improve the surface life underneath zinc coating and act as a barrier against zinc reaction with environmental species. Pinholes, air bubbles, craters, and cracks found in the paint are sealed by corroded zinc products. Additionally, zinc products that were corroded generally occupy 25% more volume than zinc, and iron oxides occupy a volume many times greater than steel; therefore, forces are reduced at the zinc-paint interface compared to those at the steel-paint interface (Revie and Uhlig, 2008). Surface degradation is a common problem in coatings when they are exposed to sunlight / UV light, moisture, pollutants, and oxygen (Nuraje et al., 2013). Therefore, by applying nano-coatings, a new approach has been found to show a better performance (Asmatulu et al., 2011). Thin-film coating technology enables control of surface-related properties, such as corrosion resistance, while allowing underlying substrates to be chosen for their strength and toughness.

Atmospheric corrosion testing of materials has a long history of application in evaluating coated and uncoated corrosion products. Salt spray test results are used to determine a material's suitability for a particular application. Salt spray testing practice is a standard method of evaluating corrosion products. For example, in galvanized steel, salt spray attacks the bare zinc. Zinc resists corrosion, but its resistance depends on the formation of reaction products. When galvanized steel is exposed to normal conditions, such as marine or road salt environment, zinc generally forms a passivation layer, mainly zinc carbonate, which resists attacks by salt and all other elements and ensures prolonged corrosion protection (Asmatulu et al., 2007; Roberge, 2008; Ahmad, 2006).

2. Experiment

2.1 Materials

Three different steel plates (G30, G60, and G90) with various galvanic coating thicknesses were selected. It was found that thicknesses higher than G90 were not only difficult to acquire but were also very uncommon according to the steel suppliers. Due to these limitations, the following steel plates were coated and tested as given in Table 1.

Table 1: Material characteristics of the steel used for this project

Zinc Coating	Steel Quality	HRB Value	Thickness
G30	CQ	40-65	0.029"
G60	DS	35-50	0.033"
G90	EDDS	15-30	0.036"

The materials selected for testing were varied in many ways than just by galvanic coating thickness as can be seen in Table 1. For example, unpainted G30 material was not available; therefore, a pre-painted sample was ordered with the understanding that the effects of corrosion may be different than the G60 and G90 samples. Additionally, the hardness of the steel varied for each sample. The "CQ" on the G30 material represents commercial quality steel, which is common in the range of 40-65 HRH (Rockwell Hardness). Similarly, "DS" on G60 material represents draw quality steel, softer steel having HRB in the range of 35-50. Finally, "EDDS" on the G90 material represents extra deep draw steel, even softer steel having HRB in the range of 15-30.

2.2 Methods

All the samples were differentiated based on zinc coating thickness. The scheduled test time in the salt spray chamber was up to 1,000 hours. Each set of samples had 3 duplicate specimens to take the average test results. The raw metal sheet materials were cut down to 7.5" x 4.0" blanks using the hydraulic shearing machine. This size was selected to support a 90° bend along the long axis and to fit in the sample trays available for the salt spray chamber. The material blanks were bent at a 90° angle on the long axis using the manual sheet metal brake. This was done to simulate a tensile load on the zinc coating on the outer edge of the bend and determine if a unique corrosion phenomenon would occur along the bend lines. A total of 12 G60 and 12 G90 samples were subjected to powder coat paint. This allowed 3 specimens for each sample to be tested at 250, 500, 750, and 1,000 hours of salt spray exposure. At the time of the procedure, the paint booth was in the process of optimizing the spray program to reduce variation in paint thicknesses. The PosiTector 6000 coating thickness gauge was used on all the painted samples after the procedure was done. The purpose of this test was to record the variation in paint thickness on the samples before testing to determine if it had any impact on corrosion formations. The salt chamber was operated in steady-state conditions at 95°F with 5% salt content (NaCl) as per the ASTM B117 standard and chemical and mechanical properties were investigated (ASTM D1654; ASTM B117). After the salt spray chamber (Singleton), the samples were cleaned in warm water to remove salt deposits. A soft brush was used to expose the corroded substrate and any visible rust on these samples. The samples were dried immediately following the rinsing and were cut furthermore using the hand lever shears to prepare for tensile strength test coupons. The tensile test coupons were extracted by a shear cutter before and after the salt spray tests.

3. Results and Discussion

3.1 Tensile Properties of Corroded Metal Sheets

All the prepared samples were subjected to tensile test to evaluate the effects of corrosion on the mechanical properties of the steel with different organic and inorganic coatings. It was initially anticipated that there would be a distinct difference in the strength of the samples as they were exposed to a salt spray chamber for a long duration. Figure 1 shows the stress-strain curve for the G60 powder-coated metal sheet samples as a function of salt spray exposure time. There was an average reduction in ultimate stress of 8% on the unpainted samples compared to painted ones, but yield stresses were nearly the same. In another case, the unpainted samples tested at 500 hours had a 21% change in yield stress over the painted variants. A 0.2% offset method was used to determine the yield strength for the G90 sample, compared to visualize and pinpoint the yield strength for the G60. A more likely possibility, however, is the fact that the thicker layer of zinc protection on the G90 allowed for less degradation in the unpainted sheets, resulting in a less severe reduction in mechanical properties. As is seen in Figure 1, the sample with 750 hours of exposure time displayed slightly better mechanical properties than its counterparts

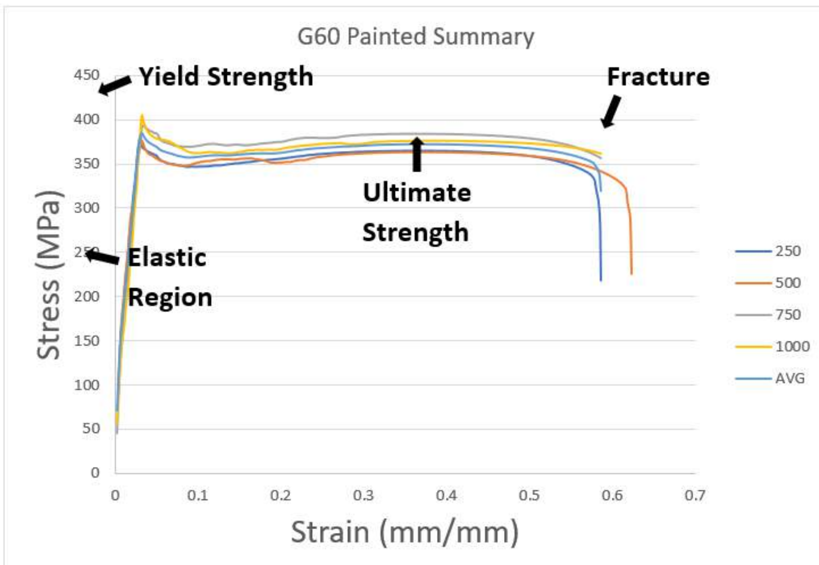


Figure 1: Stress-strain curve of the G60 powder-coated metal sheet samples as a function of salt spray exposure times

Figure 2 shows a summary of the 12-powder coated G90 samples. This material had impressive repeatability in tensile strength results across all samples compared to the G60 material. There was practically no difference in tensile strength between any of the tested durations, which may be because of the shallow surface damages and paint degradations.

Figure 3 shows a summary of the 15 pre-painted G30 samples. Like the G90 material, these samples performed very close to one another with little to no tangible impact on the structural integrity of the steel.

A summary of the 11 unpainted G60 metal sheet samples is shown in Figure 4 as a comparison. First, the samples that were untested in the salt spray chamber withstood slightly lower stress than the other samples in the beginning, but in turn managed to withstand more elongation, as well. Additionally, the 250-hour sample failed quite early compared to the rest. Most likely, these are related to the structural changes in the selected materials along with the salt spray effects.

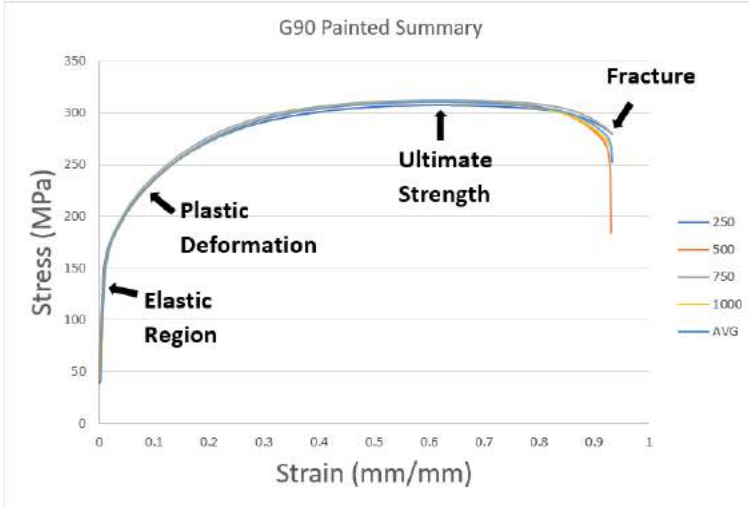


Figure 2: Stress-strain curve for the G90 powder-coated sheet samples as a function of salt spray exposure time

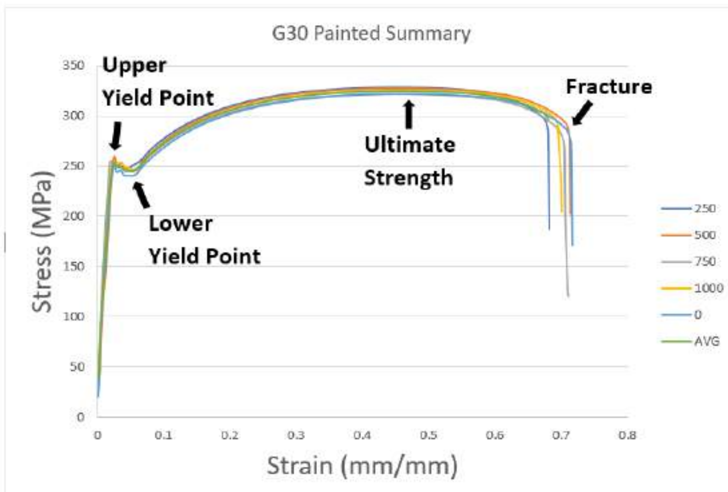


Figure 3: Stress-strain curve for G30 pre-painted samples as a function of the salt spray exposure times

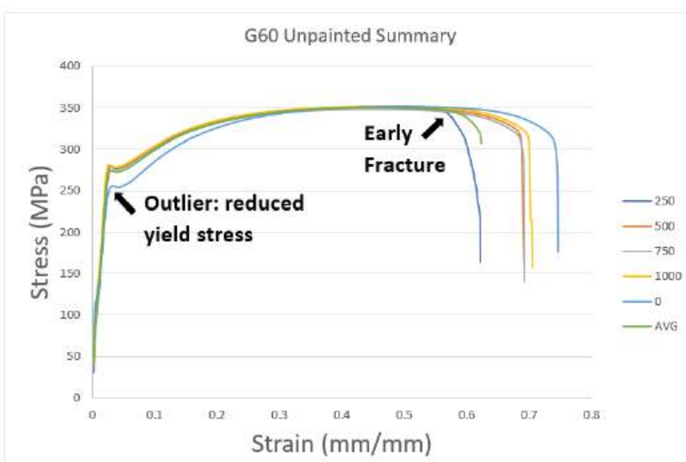


Figure 4: Stress-strain curve for the G60 unpainted samples as a function of exposure times

In these studies, about 11 unpainted G90 samples were compared to each other, and the test results are shown in Figure 5. Although the samples here overlapped nicely for the most part, particularly in terms of strain values, however, there appears to be an outlier in the 500-hour sample, like the outlier seen in the G60 comparison from Figure 4. The similar materials' property effects and the salt spray degradations.

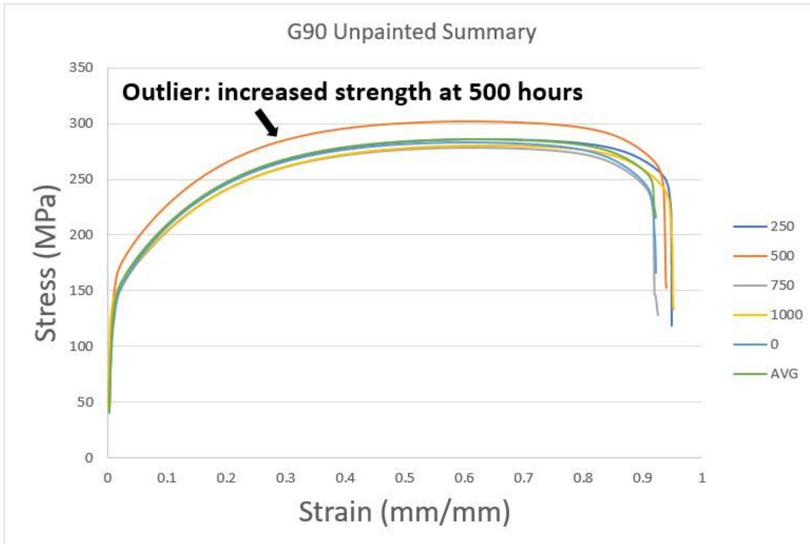


Figure 5: Stress-strain curve for the G90 unpainted samples as a function of exposure times

Figure 6 shows the comparisons among the three unpainted G60 samples. These specimens were unique in that they consisted of both 1,000-hour G60 bottom edge samples compared against a similarly sized untested G60 sample. The purpose of this comparison was to assess the most severe corrosion and rusting that may be observed throughout the salt spray chamber studies. In this case, a noticeable reduction in mechanical strength can be seen on the selected test coupons after exposure.

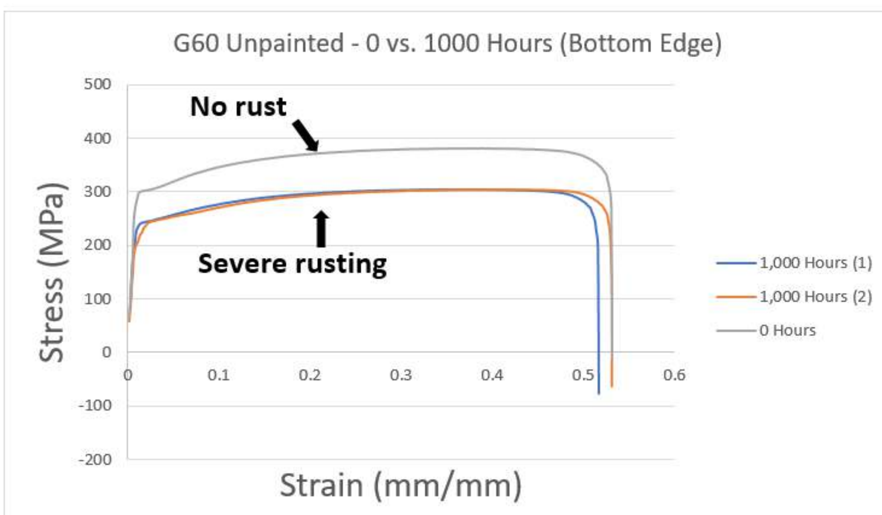


Figure 6: Stress-strain curve for the G60 unpainted samples with severe corrosion effects under the salt spray tests

Figure 7 reveals the comparisons among the three unpainted G90 samples. These specimens provided similar test results that are observed in the tests of the G60 samples. Although the difference in material strength between the severely corroded samples at 1,000 hours and the 0-hour sample was to be expected, the range of strain across all the values raises the question of which one(s) are the outliers. Based on our observation, multiple parameters can be seen on the prepared samples for the various mechanical property changes (stress, strain, elongation, elastic modulus, yield, and ultimate tensile strengths, etc.), which are mostly related to the material's structure, coating, and properties, testing conditions, local degradations, and some other unknown effects.

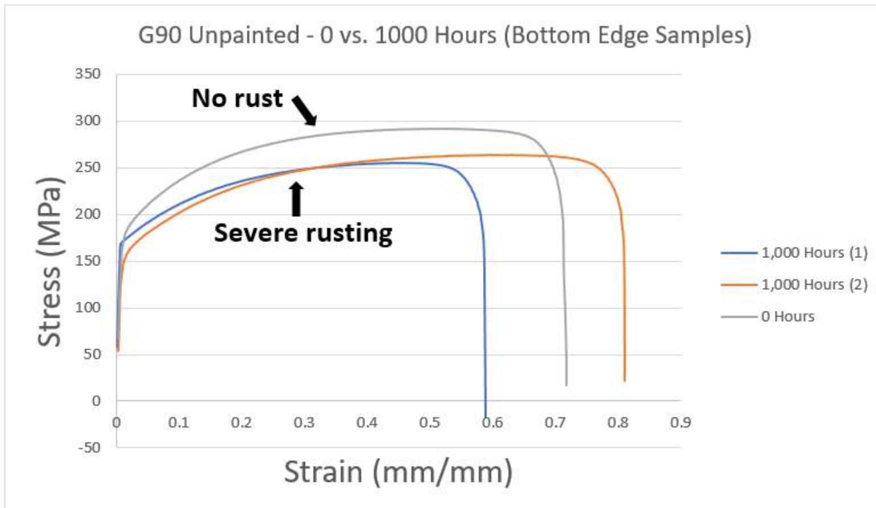


Figure 7: Stress-strain curve for G90 unpainted samples with severe corrosion effects

Tsyrulnyk and Heneha (2004) studied the effects of stressed steel and concluded an intensified dissolution of sacrificial anodes on steel in loaded or stressed conditions. The tensile strength testing provided insight into an evaluation method less commonly found in this type of corrosion testing, yet still applicable to make comparisons between different samples. Tensile strength testing is also applicable in fatigue testing. Coni et al. (2009) studied the influence of the coating on the mechanical properties of galvanized steel. They determined that the specific yield strength and the specific tensile strength of the stripped samples were higher than that of the coated samples, which indicates that the presence of coating decreases the specific yield and tensile strength of the galvanized steel sheet (Coni et al., 2009). The zinc coating displayed a higher influence on the reduction of the specific yield strength between coated sheets and striped sheets. X. Shi et al. (2009) used homogeneous epoxy coatings containing nanoparticles of SiO_2, Zn, Fe_2O_3, and halloysite clay on steel substrates to provide extra protection against corrosion and enhance mechanical properties. They determined that the mechanical properties of nanocomposites largely depend on the integrity and internal properties of the coating surface, since under mechanical stress the micro-voids between the nanoparticles or between the polymer matrix and the nanoparticles may become the origin of cracks. They also reported an increase of around 30% in young's modulus for the nano-Zn-modified epoxy coating. Structural and technological effects in some materials can change the overall properties of the sample, which causes variations in the physical and chemical properties (Subeshan et al., 2022; Zhang et al., 2013; Khadak et al., 2021; Subeshan et al., 2020; Alarifi et al., 2019).

3.2 Visual Analysis of Corroded Surfaces

All the G30, G60, and G90 samples were initially free of rust regardless of the paint finish (pre-paint vs. powder coat) or test time. The powder coating proved to be an effective and impenetrable protective coating due to its hard and smooth surface. These samples specifically had little or no salt deposits on the surface when being removed from the chamber. Similarly, the pre-painted G30 material did not display any signs of rust despite the slightly more textured surface, when compared to the powder-coated finish. The 90° bent the samples did not appear to promote any signs of degradation, as well. The coated samples have a uniform thickness of paint, and this coating would cover any of the microcracks in the zinc coating.

Although there was no red color rust seen on any of the painted samples, many samples suffered from blistering due to the corrosive effect (cathodic reactions with hydrogen gas formation). There was a noticeable difference in the blistering between pre-painted G30 steel and G60 or G90 powder-coated steel. In the case of G30, blistering occurred frequently both around the scribe line as well as away from it. This was likely due to a combination of the relatively thin layer of zinc protection (or damaged zinc film) coupled with the pre-paint finish that was less protective than the harder powder coat alternative. Figure 8 shows the blistering effects on the 750 hours and 1000-hour tested G30 samples. Blistering mainly forms near the damaged areas and on the edges.

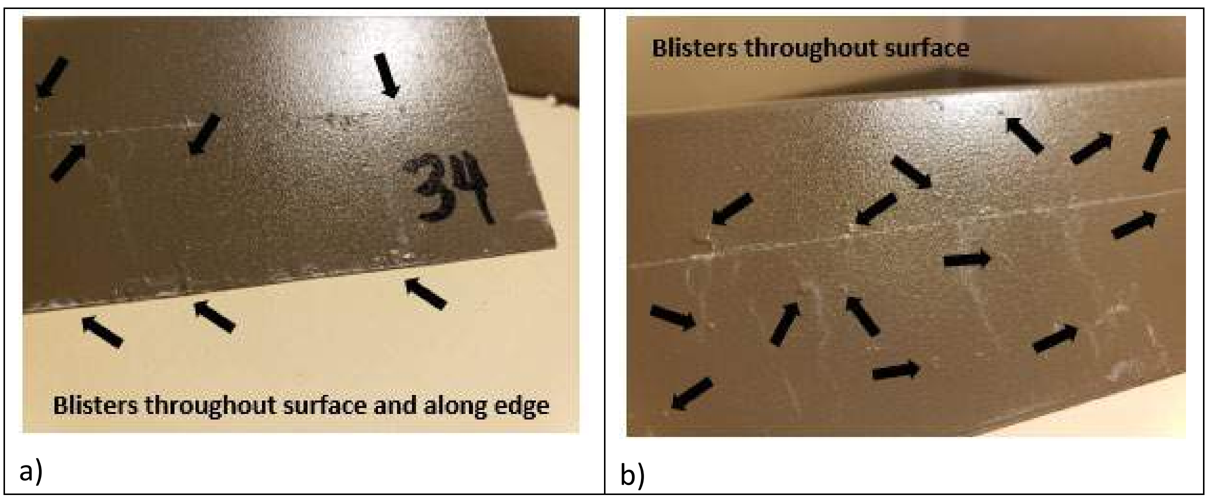


Figure 8: Blistering effects on the G30 samples for a) 750 hours and b) 1,000 hours

For the 250 and 500 hours of exposure, the samples showed some little blistering along the scribe line. The scribe line itself did not exhibit any severe corrosion outside of small traces of salt deposit embedded in the scribe. The blistering that occurred sporadically across the surface would quickly prove detrimental to the long-term reliability and appearance of the material. The G60 and G90 material that was powder coated behaved similarly to the G30 samples in some ways. The primary similarity was the blistering effects; however, it was limited to the scribe line in nearly all cases. Additionally, both the size and frequency of blisters along the scribe line were increased with these samples, as well as compared to G30. Figure 9 shows the G60 and G90 powder-coated samples after 1,000 hours of salt spray exposure times.

Additionally, some blister formations have been seen in some of the samples due to the aggressive corrosive atmosphere. Salt spray caused oxidation and destroyed the coating structures to improve the corrosion rates and reduce the mechanical properties. Also, it was reported hydrophobic and superhydrophobic coatings greatly improve the corrosion resistance of the surface by eliminating the corrosion formations and keeping the mechanical strengths of the substrates (Asmatulu et al., 2007;

Nuraje et al., 2013; Asmatulu et al., 2016; Jurak et al., 2020). Marin et al. (2011) improved the corrosion resistance of AISI 316 stainless by coating using atomic layer deposition (ALD) of Al_2O_3/TiO_2 layers.

a) b)

Figure 9: Blistering effects after 1,000 salt spray tests on a) G60 and b) G90 samples

Leppañiemi et al. (2018) used atomic layer deposition coatings to seal pinholes resulting in physical vapor deposition coatings to provide extra resistance against corrosion. Joseph et al. (2021) conducted several experiments on galvanized steel sheets for roof building and construction in rainwater and acidic environment and found that the green sheets were the least affected in rainwater with a lower corrosion rate. Additionally, the lowest corrosion rate values were observed in the rainwater experiments compared with the acidic environment tests. In hydrochloric acid (HCl) solution weight-loss experiments, the green roofing sheet showed the maximum corrosion rate compared to the cream roofing sheet. Recent studies also indicated that the corrosion properties of many metallic and organic surfaces can be altered using hydrophobic and superhydrophobic coatings and nanocomposite coatings, and the lifetime of those substrates was substantially improved. In corrosion and coating-related issues, polymer degradation is also a major issue to consider for better corrosion mitigation (Subeshan et al., 2022; Zhang et al., 2013; Khadak et al., 2021; Subeshan et al., 2020; Tran et al., 2018; Asmatulu et al., 2019).

4. Conclusions

Many steel sheets with different zinc coating thicknesses and paint finishes used for HVAC systems were tested under different salt spray conditions to evaluate degradation levels and mechanical properties. Duplex systems involving combinations of inorganic metallic-based coating systems with organic polymer-based paint finishes are commonly used in manufacturing processes, but there is a paucity of literature on this subject. Duplex coating systems experienced no (or very little) degradation of strength regardless of the number of hours exposed to the salt spray chamber environment up to 1,000 hours. This is due to the high level of protection provided by the outer organic powder coating layer and further protection from the underlying zinc layer. This was true for all cases including the G30, G60, and G90 steel sheets used for the HVAC systems. There was some decrease in the yield strength of certain unpainted steel sheets compared to painted counterparts. The sample G60 experienced a 30% reduction in yield strength due to the uniform surface corrosion, and only a small 6% reduction in ultimate tensile strength. Additionally, this same relationship was not seen between painted and unpainted G90 material and is likely due to the increased coating thicknesses.

References

Ahmad, Z. (2006). Principles of Corrosion Engineering and Corrosion Control. IChemE, New York, 2006.

Asmatulu, R., Claus, R.O., Mecham, J.B., and Corcoran, S.G. (2007). Nanotechnology-Associated Coatings for Aircrafts. *Materials Science*, 43, 415-422. https://link.springer.com/article/10.1007/s11003-007-0047-7

Asmatulu, R., Diouf, D., Moniruddin, M., and Nuraje, N. (2016). Enhanced Anti weathering Nanocomposite Coatings with Silanized Graphene Nanomaterials. *International Journal of Engineering Research and Applications*, 6, 79-91. https://www.ijera.com/papers/Vol6_issue6/Part%20-%203/E0606032336.pdf

Asmatulu, R., Erukala, K.S., Shinde, M., Gorji, M.R., and Alarifi, I.M. (2019). Investigating the Effects of Surface Treatments on Adhesion Properties of Protective Coatings on Carbon Fiber-Reinforced Composite Laminates. *Surface and Coatings Technology*, 380, 125006. https://www.sciencedirect.com/science/article/abs/pii/S0257897219309971?via%3Dihub

Asmatulu, R., Mahmud, G.A., Hille, C., and Misak, H.E. (2011). Effects of UV Degradation on Surface Hydrophobicity, Crack and Thickness of MWCNT-Based Nanocomposite Coatings. *Progress in Organic Coatings*, 72, 553-561. DOI:10.1016/j.porgcoat.2011.06.015

ASTM B117. (2018). Standard Practice for Operating Salt Spray (Fog) Apparatus. Retrieved January 21, 2019. https://www.astm.org/b0117-16.html

ASTM D1654. (2016). Standard Test Method for Evaluation of Painted or Coated Specimens Subjected to Corrosive Environments. January 21, 2019. https://www.wewontech.com/testing-standards/190306003.pdf

Bhadu, M. K., Guin, A. K., Singh, V. and Choudhary, S. K. (2013). Corrosion Study of Powder-Coated Galvanised Steel*. ISRN Corrosion*.. https://www.hindawi.com/journals/isrn/2013/464710/

Coni, N., Gipiela, M. L., D'Oliveria, A. M. P., and Marcondes, V. P. (2009). Study of the Mechanical Properties of the Hot Dip Galvanized Steel and Galvalume*. Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 31(4), 319-326. https://www.scielo.br/j/jbsmse/a/NcWxhsMGgmr3LphpsRZgVkf/s

Guin, A. K., Nayak, S., Rout, T. K., Bandyopadhyay, N. and Sengupta, D. K. (2011). Corrosion resistance nano-hybrid sol-gel coating on steel sheet. *ISIJ International*, 51, 435–440. http://publicationslist.org/data/Suryakanta_Nayak/ref-1/ISIJ%20International,%20Vol.%2051%20(2011),%20No.%203,%20pp.%20435%E2%80%93440.pdf

Hernandez-Alvarado, L.A., Hernandez, L.S. and Rodriguez-Reyna, S.L. (2012). Evaluation of Corrosion Behavior of Galvanized Steel Treated with Conventional Conversion Coatings and a Chromate-Free Organic Inhibitor. *International Journal of Corrosion*. https://doi.org/10.1155/2012/368130

Joseph, O.O., Dirisu, J.O., and Odedji, A.E. (2021). Corrosion resistance of galvanized roofing sheets in acidic and rainwater environments. *Heliyon*, 7(12). https://www.cell.com/heliyon/pdf/S2405-8440(21)02750-X.pdf

Jurak, S., Jurak, E., Uddin, M. N., and Asmatulu, R. (2020). Functional Superhydrophobic Coating Systems for a Possible Corrosion Mitigation. *International Journal of Automation Technology*, 14, 148-158. DOI:10.20965/ijat.2020.p0148

Khadak, A., Subeshan, B., and Asmatulu, R. (2021). Studies on De-Icing and Anti-Icing of Carbon Fiber-Reinforced Composites for Aircraft Surfaces using Commercial Multifunctional Permanent Superhydrophobic Coatings. *Journal of Materials Science*, 56, 3078-3094. https://doi.org/10.1007/s10853-020-05459-9

Leppañiemi, J., Sippola, P., Peltonen, A., Aromaa, J., Lipsanen, H. and Koskinen, J. (2018). Effect of Surface Wear on Corrosion Protection of Steel by CrN Coatings Sealed with Atomic Layer Deposition. *ACS Omega*, 3, 1791-1800. https://pubs.acs.org/doi/pdf/10.1021/acsomega.7b01382

Marin, E., Lanzutti, A., Guzman, L. and Fedrizzi, L. (2011). Corrosion protection of AISI 316 stainless steel by ALD alumina/titania nanometric coatings. *Journal of Coatings Technology and Research*, 8(5), 655-659. DOI:10.1007/s11998-011-9327-0

Nuraje, N., Khan, S.I., Misak, H.E., and Asmatulu, R. (2013). The Addition of Graphene to Polymer Coatings for Improved Weathering. *ISRN Polymer Science*, 2013. https://doi.org/10.1155/2013/514617

Revie, R.W. and Uhlig, H.H. (2008). Corrosion and Corrosion Control –An Introduction to Corrosion Science and Engineering (4th ed.). Wiley-Interscience, New York.

Roberge, P. R. (2008). Corrosion Engineering Principles and Practice. McGraw Hill, New York. www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1663750

Subeshan, B., and Asmatulu, R. (2022). Corrosion Mitigation of Metals and Alloys via Superhydrophobic Coatings with Plasma Surface and Heat Treatment Processes. *Engineering Failure Analysis*, 139, 106437. https://doi.org/10.1016/j.engfailanal.2022.106437

Subeshan, B., Usta, A., and Asmatulu, R. (2020). Deicing and Self-Cleaning of Plasma-Treated Superhydrophobic Coatings on the Surface of Aluminum Alloy Sheets. *Surfaces and Interfaces*, 18, 100429. https://doi.org/10.1016/j.surfin.2020.100429

Tran, V. T., Abedin, F., Usta, A., and Asmatulu, R. (2018). Polyurethane Nanocomposite Coating with Silanized Graphene and Hexagonal Boron Nitride as Nanoadditives for Improved Resistance against UV Degradation. *Journal of Composite Materials*, 53, 1387-1399. https://www.deepdyve.com/lp/sage/polyurethane-nanocomposite-coating-with-silanized-graphene-and-lH0xt1vqHb

Tsyrulnyk, O.T. & Heneha, B. Y. (2004). Mechanisms of the Sacrificial Protection of Stressed Low-Alloy Steel. *Materials Science*, 40(5), 597-604. https://link.springer.com/article/10.1007/s11003-005-0087-9

Shi, X., Nguyen, T. A., Suo, Z., Liu, Y., Avci, R. (2009). Effect of nanoparticles on the anticorrosion and mechanical properties of epoxy coating. *Surface and Coating Technology*, 204, 237-245. DOI:10.1016/j.surfcoat.2009.06.048

Zhang, B., Patlolla, V.R., Chiao, D., Kalla, D.K., Misak, H., and Asmatulu, R. (2013). Galvanic Corrosion of Al/Cu Meshes with Carbon Fibers and Graphene and ITO-Based Nanocomposite Coatings as Alternative Approaches for Lightning Strikes. *International Journal of Advanced Manufacturing Technology*, 67, 1317-1323. https://link.springer.com/article/10.1007/s00170-012-4568-3

Challenges to the Implementation of Lean Practices in Higher Education: Is the Culture in Higher Education Conducive to Adoption of Lean Principles and Philosophy?

Christopher Kluse[1*]
Chris J. Shannon[2]
[1]*Bowling Green State University, USA*
[2]*University of Queensland, Australia*
ckluse@bgsu.edu

Abstract

Lean in higher education, often abbreviated as LHE, has grown in interest over the past 20 years. The application of lean principles has been observed in higher education institutions throughout the world. Many researchers have investigated both the success and the failure of lean in higher education. Although lean originated in the automotive industry, and both the automotive industry and higher education report success and failure, there is no comparison of the factors between these groups. Culture is cited quite frequently in the literature as a barrier to success. Thus, there is a need for investigation into the cultural differences between higher education and the automotive industry as they relate to lean. This paper uses Schein's three elements of culture to evaluate the differences in culture between higher education and the automotive industry. It concludes with recommendations on how to best deploy the practices of lean in higher education given the cultural differences and challenges between the two sectors.

Keywords: Lean, culture, higher education, automotive industry, levels of culture.

1. Introduction – Lean success and failure

Higher education in Western countries has experienced many challenges in recent decades including a reduction in public and government funding, increased competition for students in a global education market, increased questioning of the value of tertiary education as compared to skilled trades or other workforce-related credentials, the academic tenure system being challenged, and most recently, the circumstances surrounding the COVID-19 pandemic. Quality management practices have been introduced and adapted into higher education for many years, however, the benefits of quality practices are still unclear and debatable (Harvey & Williams, 2010). Lean practices in Higher Education have increased in the past two decades, and, similar to applying quality management practices, the benefits from applying lean concepts in Higher Education are not clear, or at least are not measured quantitatively. For example, Balzer et al. (2016) conducted a review of 64 publications between 2009 and 2015. These publications consisted of case studies, (52%), conceptual papers (32%), and empirical research (9%). While the case studies were mostly positive towards the effective use of lean principles in higher education, empirical studies failed to measure the effectiveness of lean in higher education. Therefore, it is still uncertain if lean practices have a positive effect on the performance of higher

Submitted June 27, 2022
Revised August 17, 2022

education institutions. On the other hand, researchers such as Richter (2011) and Scoggin (2017) estimate the failure rate of lean in higher education anywhere from 50% to 95%, indicating lean initiatives often fail. Specific examples of this include Cano, Reilly, and Kouroukli s (2014) who present a case study that targeted the improvement of administrative and academic processes using lean practices at a higher education institution in Scotland. The study concluded that the lean initiative failed to achieve the intended benefits and Shannon (2020) described lean as having a positive impact in a discrete unit while the university-wide implementation failed. In general, many researchers have investigated both the success and failure of lean in higher education. Factors that have been identified as having an adverse effect on the success of lean initiatives include, but are not limited to, leadership support, organizational culture, poor project selection, use of temporary lean experts, disjointed lean strategy vs organizational strategy, and the discrete use of lean tools vs an organizational-wide holistic implementation of the lean philosophy.

Lean originated in the automotive industry, and while both the automotive industry and higher education report success and failure, there is no comparison of the organizational and cultural factors between these sectors. As a result, this paper seeks to explore and identify these factors and their effect on the success of lean within the higher education sector. The authors seek to understand if lean practices are transferable given the cultural and operating differences between the automotive manufacturing sector and higher education; and, if these factors inherently are barriers to the successful application of lean in higher education, what can be done in the future to best exploit lean in the HE environment? Culture is cited quite frequently in the literature as a barrier to success, in both manufacturing and higher education, thus, an examination of the cultural differences may lead to an understanding of whether lean can be transferred to higher education or if there is a natural constraint based on culture. This paper focuses on the difference in culture between the higher education community in North America and Australia, and the automotive community in North America. It concludes with recommendations on how to best exploit the practices of lean in higher education given the cultural differences and challenges between the two sectors.

2. Problem statement

Lean principles have been applied across the globe in the higher education sector both with success and failure. There is a need to understand if, and why, cultural barriers exist in higher education that may limit the potential value of applying lean practices in a higher education environment.

3. Research questions

Q1. Do operating differences and norms between higher education and the automotive manufacturing sector pose a barrier to successful lean initiatives in higher education?

Q2. How can the lean philosophy be adapted to suit the needs of contemporary HE?

4. Method

The three levels of culture as outlined by Schein (2017), a comparison of the artifacts, espoused beliefs, and underlying assumptions of organizations, are used to assess the cultural differences between Higher Education and the Automotive industry. According to Schein, “culture can be analyzed at several different levels, with the term “level” meaning the degree to which the cultural phenomenon is visible to you as a participant” (Schein, 2017, p. 17). Schein asserts that each of these levels can range from the discrete and tangible level of culture to the subtle, unconscious, and inherently entrenched aspects that define a particular culture. Using the three levels of cultural analysis, the researchers present (in table 1) observations from the automotive industry and higher education that in turn manifest as key differences between the two sectors. Pondering why these differences exist then lead the researchers to break down each dimension of the respective industries, outline the accepted underlying assumptions (for the researchers combined experience in these 2 industries), and provide examples of tangible artifacts that support the underlying assumptions. It is from this framework and analysis, the potential for lean to succeed or fail within higher education can be objectively evaluated. To help better understand these levels of culture, a brief description of each level is outlined below.

4.1 Authors Experience

These authors intend to introduce this important topic to the higher education community since lean is being applied at colleges and universities worldwide. Due to limited published work in this area, personal experience and dialog with colleagues form the foundation of this research. The researchers bring their collective experience in higher education and Kluse's experience working in quality management roles in the automotive industry, along with input from their respective networks, to help inform this paper. Before a full-time academic appointment in 2015, Kluse was a former Quality Professional with the majority of his 27-year industry career spent in the automotive sector and working in a lean environment with automotive component suppliers and OEMs. Shannon brings over 20 years of experience in the higher education sector working in central divisions, faculties, schools, and research centers.

4.2 Artifacts, espoused beliefs, and assumptions

Artifacts, according to Schein (2017), are:

> the phenomena that you would see, hear and feel when you encounter a new group with an unfamiliar culture. Artifacts include the visible products of the group, such as the architecture of its physical environment; its language its technology and products; its artistic creations; its style, as embodied in clothing, manners of address, and emotional displays; its myths and stories told about its organization; its published list of values; and its observable rituals and ceremonies" (p. 29).

As such, most of what is visible in an organization is a form of artifact.

Espoused beliefs and values are those beliefs and values that members of the organization say they believe in and desire. These can be reflected in missions and value statements, included in statements from senior executives, be included in the content on websites and in promotional material, and so on. They are the things that the organization claims to be, or would like to be.

Underlying assumptions "are things that are taken for granted as being reality in the organization" (Schein, 2017, p. 30). Unlike espoused beliefs and values, these assumptions are embedded in the actions and behaviors of the organization and inform the way things are done

The challenge in understanding a different culture, be it an organizational culture or a national one, is that artifacts can look similar across cultures, but have very different meanings and uses. As Schein (2017) puts it, "It is especially dangerous to try to infer the deeper assumptions from artifacts alone, because your interpretations will inevitably be projections of your own cultural background" (p. 29).

5 Results

Using the framework described in the previous section, common aspects of lean in the automotive and higher education sectors have been identified and further explored in table 1 and table 2. It is from this analysis the potential for lean to succeed in higher education can be assessed and the future direction of lean in higher education determined.

Table 1 lists different aspects of lean implementation and the way that the underlying assumptions manifest during implementation.

Table 2 lists a number of characteristics that exist in both the auto industry and higher education, followed by statements of the underlying assumptions and examples of the artifacts in each industry. For example, quality assurance is essential for the automotive industry to sell cars. Because there is only a need to accredit some degree programs and not all aspects of university business, quality assurance is seen as optional.

6. Analysis

After analyzing common aspects of Lean in terms of artifacts, assumptions, and espoused beliefs, it is clear that cultural differences exist between the two sectors, and these surely contribute to the effective application of Lean in Higher Education. In the second edition of The Toyota Way, Dr. Jeffrey Liker (2021) presents two differing strategies for the implementation of lean. Liker initially presents what he terms "*a comfortable tendency toward mechanistic implementation (of lean)*" which is simply a non-preferred method to initiate a lean conversion within the organization. In brief, this entails the development of lean metrics that align with lean principles, the metrics are tied to management bonuses to spark motivation and commitment. An initial project is executed using the kaizen approach (aka as rapid improvement events or RIE in HE) with success; this strategy is taken forward through the organization by conducting monthly kaizen events, for several consecutive months – each led by a manager or engineer at each facility. Results are exceptional; however, they are not sustained. This is attributed to the management and/or staff not taking ownership. Without going into further detail about the mainly failed initiative, one can draw comparisons of this scenario to lean in higher education. An administrative process is chosen to improve, ownership and buy-in may or may not be established, and results *are* realized, however no real lean initiative or lean conversion takes place in the institution even when multiple kaizens are successfully conducted over a period of time. It implores the question – are we in HE culturally sound to accept what lean has to offer? Alternatively, is our cultural make-up and foundation counterintuitive to lean principles even if they are generally thought of as simplistic (reduce waste in all processes) and well aligned with most any industry?

Liker (2021) goes on to present a better approach to lean transformation whereby one is “Approaching Lean Transformation Scientifically” (p. 355). In this approach, before the organization delves into solutions (i.e., initiating periodic kaizen events), a big-picture look is established where the organization ponders questions based on an improvement model. According to Liker, these questions are: “(1) what is the challenge? (2) what is the current condition? (3) what is the next target condition, and what are the obstacles to that condition?, and (4) what is the next experiment I will run to overcome an obstacle?” (p. 355). Without presenting the entire approach in great detail, it is from this approach that lean conversion is most likely to succeed and results sustained. In presenting these two dissimilar strategies to lean implementation, Liker ultimately defers to culture as the ultimate building block for sustained lean practices. According to Liker (2021) “It seems that whatever the starting point of discussion of lean transformation, we end up talking about culture. Perhaps this is an indication that

Table 1. Common Aspect of Lean – Perspective from Each Sector

Lean Perspective	Auto Industry Response	Higher Education Response
Lean acceptance	Lean is (essentially) universally accepted in the auto industry – best practices revolve around lean and/or TPS. Awards are lean-based (AME, Shingo, etc.)	Lean viewed as fad or latest initiative of the year, acceptance by HE unclear, and/or no universal acceptance
Lean rationale - to use lean techniques	Lean is applied to the core business of making automobiles. It is assumed that lean practices are required to compete with other OEMs	Lean is rarely applied to the core business of teaching and research. It is mostly applied to administrative processes that are deemed broken
Sense of urgency (to correct inefficient or poor performing processes)	Efficient processes and defect-free products are considered a must for competitive advantage and to sustain low-profit margins and/or avoid costly OEM production delays, line stoppage, or stock-outs	Efficient processes desired, urgency not necessarily important for competitive advantage
Elements of lean in use	Lean was created in auto industry. All elements are logically applicable to manufacturing	Some lean practices such as JIT or TPM are not well aligned with university processes or functions
Leadership/Administration	Industry leaders are often evaluated by discrete organizational metrics such as profitability, quality, productivity, etc., and rewarded for performance. Leaders are typically trained in leadership, management, and process improvement principles Leadership can form a team, they have direct responsibility for team performance, create a vision and execute a strategy. Concept of Gemba "walks" practiced in industry. Leaders have direct authority over subordinates	University leadership is often senior academics that have excelled in a particular area of professorship such as externally funded research. Leadership often do not have formal training in management or leadership principles Leadership in HE supervise faculty that are generally driven by individual goals tied to rank and tenure. Gemba "walks" rare – processes solved based on intuition. Leaders in academia do not necessarily have influence over subordinates
Leadership focus	Lean is applied to the core business of making automobiles and is critical to organizational success Leaders/managers within the auto industry are generally versed in process improvement techniques and this is part of their daily responsibility – to monitor and improve all organizational processes	Lean is rarely applied to the university's core business of teaching and research and is seen as tangential to organizational success Leaders/administrators in HE do not have daily focus on processes or process metrics, and are not trained in process improvement techniques
Perception of the workforce	The workforce is treated as "process experts" and often the first consulted when a problem arises. Leadership's responsibility is to remove barriers to problems, workforce solves problems	Leaders (administration) solve problems or lead problem-solving efforts, especially in academic processes. The actual workers in the process and/or faculty are not always viewed as the process expert
Vision and Mission	Vision and mission are often quite clear and concise. The ultimate goal for the auto industry is to maximize profit, satisfy stakeholders, and continued business, and growth. Lean practices are a significant aspect that supports and help execute a company's vision and mission. Short and long-term strategies based on the exploitation of lean principles, considered a "must" for competitive survival	The University's mission and vision are driven by the board of directors, public need, workforce need, and administrators. Vision and mission holistic. The ultimate goal for HE can be somewhat varied, it can be the number of graduates, research funds obtained, university ranking, athletic success, or efficient use of tuition. Lean principles do not align with the vision and mission

Table 2. Cultural Analysis in Automotive Sector vs. Higher Education

Characteristic/Dimension	**Auto Industry Assumption & Artifacts**	**Higher Education Assumption & Artifacts**
Team focus	Teams are considered a necessity and organizational success is equated to high-performing teams. Teams-based activities are essential for success **Artifacts:** Program launch teams, problem-solving teams, design teams, etc. AIAG core tools suggest the use of "cross-functional" teams. IATF 16949 requires the use of multi-disciplinary teams	HE is focused on individual academics. Individual success awarded/promoted by internal and external practices. Academics collaborate as needed in self-forming networks **Artifacts:** Tenure track, personal promotion, indices like H Index, Fellowships.
Financial expectations	Low-profit margins demand efficiency. Mandated annual program "give backs" (LTA's) or pay-to-bid on a program force the need to consistently reduce costs/waste and improve through efficiency gains. Customer price considered "fixed" in lean **Artifacts:** OEM LTA's, OEM supplier cost audits,	Internal and external expectations that costs rise each year. Pass along cost to "customer" i.e. students. Cost is variable. Universities seek to establish large endowments via philanthropy. **Artifacts:** Enterprise bargaining increases, indexed grant funding, annual fee increases, tuition increases
Requirement for quality assurance	Required & suggested lean practices outlined in IATF 16949 certification and various other industry standards such as AIAG core tools **Artifacts:** IATF registration required by all suppliers to OEM. Within the standard, improvement is mandatory, use of lean encouraged/ required. Quality can be discretely defined and measured. Poor quality results in potential loss of certification and business	Formal accreditation only applies to some degree programs. There is no requirement for specific practices to improve support and/or core processes **Artifacts:** Accreditation processes for particular programs only. Focus on achievement of learning outcomes, cost and efficiency are not considered. Quality is not easily defined
Final product	Makes and sells a tangible product – motor vehicles. Quality is standardized and readily measured/assessed with common metrics across the industry **Artifacts:** Public perception of quality, JD Power surveys that dominate OEM actions and leadership decisions. Specific measures reported common to each OEM, quality assessed by OEM and during third party certification & surveillance audits	Universities make and sell intangible products – education and research. Assessing quality is subjective and not standardized **Artifacts:** Institution, state, and national teaching quality awards, student evaluations of courses, institution, and national research excellence measures and awards
Competition	Clear understanding of performance relative to competitors through internal analysis, industry reporting of metrics, and competitive benchmarking **Artifacts:** Weekly production reports in "auto news", accepted productivity benchmarks among the industry, JD Power quality rankings, etc.	Universities have broad understanding of relative status via externally managed rankings. Each ranking system has its own measures of success **Artifacts:** University ranking such as ARWU, THE, QS World University
Definition of the customer	The customer in the automotive sector is clearly defined **Artifact:** For an automotive component supplier the customers are the subsequent operation during assembly, the OEM, and the consumer. For the OEM, customers are the subsequent operation during assembly and the consumer.	The customers for the University are often debated and not clearly defined **Artifact:** Students, the public (service and consulting), and industry (applied research) are all measured and valued differently. Administration, faculty, and staff have differing end customers.

culture is at the root of everything I have been discussing" (pg 364). Furthermore, Liker goes on to cite Schein and his definition of culture resulting in the following conclusion: "This [Shein's definition of culture] is a remarkably apt description of the Toyota Way culture in a number of ways:" (p. 365). Liker goes on to describe specific aspects of Toyota's culture that directly align with the definition of culture and then offers a model for a successful lean journey. Examination of the model shows a clear disconnect between the vision, mission, and culture in higher education with that of a successful lean conversion.

6.1 Main themes – Cultural analysis

Analysis of table 1, aspects of lean as viewed from the automotive industry and higher education perspective, indicates that many items in the table are not well aligned to cultural characteristics conducive to a healthy lean atmosphere. Regarding ***acceptance of lean, leadership, and vision/mission***, in general, lean is sporadically accepted in higher education and may also be only partially accepted within a university that seemingly practices lean. In contrast, universal acceptance is a norm in the automotive industry. Since culture is based on assumptions and shared beliefs, attempting to implement lean principles without a culture conducive to lean, and explicitly modeled by leadership is a clear barrier to lean. It is not a belittling statement of higher education culture, but it represents a situation that becomes quite important if lean is to be successful in higher education. Liker's (2021) model suggests that at the start of a lean journey, leadership vision and commitment to excellence are critical; without this commitment, the next logical path is to attempt to exploit lean tools resulting in short-term results that are not sustained (p. 369). Therefore, institutions that wish to successfully implement lean need to work on their culture before running any kaizen events or 'lean awareness' training.

Looking at the ***sense of urgency*** between the 2 sectors, it is evident that the interest in lean within higher education represents more of a suggested improvement approach rather than an immediate need to correct poor-performing processes. While improving processes is surely important in higher education, it does not have the same motivation as in the automotive industry where profit maximization is the ultimate goal. In contrast, the motivation for process efficiency in HE does not carry the same implications (degradation of profits) as those in the automotive sector.

It can be observed that many ***lean elements***, concepts, and tools such as Just in Time (JIT), Total Productive Maintenance (TPM), Single Minute Exchange of Die (SMED), involvement of people, Jidoka – just to cite a few examples, do not align with university operations and processes. It is not to say that some can be applicable or transferrable to some university operations, however as previously stated, lean taken as a holistic endeavor, rather than the application of selected tools, is the preferred and most conducive to success approach (Liker, 2021).

Regarding ***leadership focus***, leaders/managers within the automotive industry are generally versed in process improvement techniques and this is part of their daily responsibility, whereas, in HE, this responsibility or focus often only surfaces when a glaring problem exists in a process. Doing something about the problem often requires bringing in external experts. The discussion points outlined in table 1 indicate that cultural differences serve as a barrier to lean in higher education.

Upon examination of ***leadership, administration, and leadership focus***, clear disconnects exist between leadership skill set, motivation, and philosophy in higher education which may not align well with the mission and vision of the university, whose focus and reason for existence (research, transfer of knowledge, public service, etc.) differs greatly from the automotive industry (profitability). The ***mission and vision*** of the university are often driven by public need, a board of directors, and university administrators while the automotive industry in the US is driven by profitability and market share. Lean principles were developed by Toyota in the 1950s, in the context of an automotive industry in Japan that exhibited: (a) a demand for a variety of products in low volume, (b) heavy competition both domestic and global, (c) fixed/falling prices, (d) dynamic technology, (e) a high cost of capital after WW2, and (f) a workforce that demanded involvement. The need for lean was a necessity for survival and this same view within the automotive industry is held today (Pascal, 2015). Therefore, the need for lean and its alignment with the university's core mission is assuredly different than that of the automotive sector.

Table 2, an examination of assumptions and artifacts, offers further insight into the cultural aspect of lean implementation. The underlying assumption that ***teams*** are a foundation of collaborative problem-solving and project execution in the automotive industry differs from the individualistic approach indicative of higher education. As observed in the assumptions and artifacts, the exploitation of teams is not only considered a preferred approach in the automotive industry; it is a preferred and implied required practice as part of achieving IATF 16949 certification. In contrast, while collaboration in higher education exists and is encouraged, formal teams are not necessarily the norm. Additionally, at the core of career advancement are individual metrics such as tenure, sole author publication, and obtaining external funding. The academic career is based on individual success, and this success is not dependent on institutional success. In fact, the success of individuals positively impacts university rankings and so drives institutional success.

In the higher education sector, it is often an assumption that ***costs*** will rise and these costs are then passed along to the customer through increased tuition fees or indexed grant funding. By contrast, in the automotive industry, it is often the expectation for suppliers to the OEMs to improve efficiency and reduce costs as a program (vehicle model) matures over time. This comes in the form of long-term agreements (LTA's) that stipulate cost reductions to the OEM by a contractually agreed upon value (i.e 3% reduction over 3 to 5 years) over the life of the program. Additionally, an emerging practice is to require an automotive component supplier to pay a fee just to get an *opportunity* to bid on a program. Thus, in the automotive industry, the use of lean principles becomes a necessity for survival. In higher education, lean can be one of many options to improve efficiency – but without the same sense of urgency or critical need.

From this analysis, the research questions are answered. In question 1, the authors inquired: *Do operating differences and norms between higher education and the automotive manufacturing sector pose a barrier to successful lean initiatives in Higher Education*? Based on the analysis and the comparisons outlined in table 1 and table 2, it can be concluded that operating norms and cultural differences do act as a constraint to the use of lean in a higher education setting. Question 2 asks: *How can the lean philosophy be adapted to suit the needs of contemporary HE?* Both tables outline these differences using the Schein (2017) model for the identification and analysis of culture. Furthermore, the researchers offer perspective, based on these differences, on what aspects of lean are well aligned for application in a higher education setting. This perspective is presented in the subsequent section.

7. Discussion

This paper discusses the fundamental cultural differences between the automotive industry and higher education. It also discusses observed responses to lean implementation in each industry. Accepting the validity of these differences, it is evident that attempts to implement lean, six sigma, or any other improvement methodology, are unlikely to succeed if treated as a *copy and paste* from manufacturing. A more nuanced and context-sensitive approach is required.

When Taiichi Ohno (2013) was leading the development of the Toyota Production System (TPS), he used a highly experimental approach and worked within the constraints of Japan's highly industrialized workforce. “Nobody knew if the Ohno System would work. Nobody else was trying it. If the results were good, that was good, and if the results were bad we needed to change right away” (p. 80). In seeking to apply lean to its own unique culture and environment, there is no reason why lean practitioners in higher education cannot also be innovative and experimental, keeping what works and changing what does not. For instance, the predominance of the kaizen event as a lean practice may well address cultural needs in higher education (e.g., employee engagement in decision-making) but only if the culture also supports the implementation of ideas from the bottom up.

7.1 Lean in academia as well as administration

A fundamental difference between lean implementation in the automotive industry and higher education is the scope of work being '*leaned*'. The two clearest differences between industries are that the automotive industry has one core business – making motor vehicles – while universities have at least two – teaching and research – which are very different from each other. The second difference is that the automotive industry uses lean on the production line in the manufacture of vehicles. That is, lean is directly related to the core business of the automotive company. Universities typically use lean on administrative processes, and not on their core business activities. As such, lean is used on hygiene factors rather than motivators. It should not be surprising that academic leadership enthusiasm for lean is limited. With that said, a better opportunity to apply lean may lie with faculty as they can choose to implement lean practices with or without the support of leadership. Emiliani (2015) states that choosing to adopt lean teaching practices without any support from leadership – even at the department level – align with the professional responsibility of the faculty member. Emiliani (2015) suggests faculty to “get on board” regardless of leadership support (or lack thereof) and the potential benefit is an “immediate positive impact on one’s students – perhaps hundreds of students per year” (p. 2040). Furthermore, Emiliani suggests that the process, starting with one faculty member using lean could subsequently spread to others via shared experience thereby slowly catching on and spreading to a group or team of faculty that embraces lean principles. As a look at where lean can go in higher education, the approach by Emiliani seems the most logical and realistic approach to implementing and sustaining lean in higher education. The approach is aligned with a core focus of the university, which is teaching, and the customer, students, thus is clearly a value-added activity.

Wiegel (2020) describes the work done at the New Engineers School as an example of lean education. Distinct from the application of lean to already established higher education institutions as described in Balzer (2020), lean education is a lean approach to teaching as well as professional support to alter an educational system that "wastes talent of students and teachers on a grand scale" (2020:vii). Wiegel describes the founding of a small engineering school running on lean principles from inception.

7.2 Local implementation rather than the whole university

The benefit of a local implementation is that it can be implemented at a lower cost and with a lower profile in the university. A discrete organization unit may also be easier to effect change in, i.e. common issues are more easily identified, local leadership may be more trusted than university-level leadership, and the intended outcomes of the lean initiative may be more readily believed. The drawback to a local implementation is that many of the processes which impact most of the staff and students run across organization unit boundaries and are owned by central divisions or functions (e.g., the student admissions process or the new staff member account setup process will be centrally owned). Therefore, effecting meaningful change may be difficult. Shannon (2020) describes success in a veterinary teaching hospital set within a university. The hospital focused on processes unique to itself and which were owned end-to-end within the hospital. They did not require the rest of the institution to change in order to see improvement. Deviating from the accepted or preferred organizational approach to implementation of lean and lean practices, a local implementation model in higher education could be much more aligned with the culture and operations of the university thus increasing the potential for success.

7.3 Simpler rather than complex

The kaizen event is the most common form of lean intervention in higher education, at least according to the available research. Instituting this simple kaizen concept locally and/or throughout the entire institution could have the snowball effect of incremental and continual improvement which is the direct intent of a kaizen program. While kaizen is considered a component of lean, taken as a single improvement philosophy in a higher education setting may prove quite valuable since it is readily aligned with the operations and culture of the university, so long as leadership will support the implementation of changes driven from the bottom up. However, there are many other forms of lean-inspired improvement initiatives which do not require the investment of time and effort that a multi-day kaizen event does. Embedding lean thinking in other activities, via existing organization development or continuous improvement teams may generate business and process improvements and also improve institutional readiness of wider programs in the future.

Balzer (2020) includes over 40 pages of examples of the successful application of lean in higher education, from 16 different universities over a 15-year period. These examples include kaizen events (referred to as rapid improvement events), process mapping, customer journey maps, structured problem solving, daily stand-ups, leadership walks, and 5S. They worked on processes related to human resources, student or academic administration, facilities, research administration, and finance among others. The universities used models ranging from small lean teams to the widespread use of lean champions across the institution. The authors are aware of other models which embed lean thinking in staff training initiatives without referring explicitly to lean, or any specific methodology. It is simply presented as useful ideas and tools to use. There have also been substantial changes to the way lean has been used in recent years due to COVID-19. The pivot to online teaching via platforms such as Zoom can also be seen in the way non-academic staff work, and the way lean practitioners work in universities (Shannon 2021, p. 47, 75-76).

8. Limitations of research

The authors acknowledge the following limitations to the research informing this paper:

While the theoretical framework to assess the cultural differences between the automotive sector and higher education is grounded in Schein's (2017) theory of organizational culture, the specific aspects of the two industries described in Tables 1 and 2 are drawn mainly from the author's experiences, observations, and discussions with colleagues. There is a lack of published literature comparing the two industries to validate the author's views.

There is a general lack of rigor in assessing the success of lean in higher education, as evidenced by the variety of approaches used, frequent use of anecdotal evidence to describe success or failure, and lack of any consistent measurement across institutions (Balzer 2015, 2019). The comparison of culture between the automotive industry and higher education as it relates to lean practices is mainly scarce.

9. Conclusions

After a preliminary evaluation using Schein's model for culture, it appears that the challenges and constraints to implementing lean in higher education are related to the difference in assumptions, beliefs, and norms associated with both higher education and the automotive sector. This does not indicate that lean will not succeed in a Higher Education setting, however, it does suggest that before launching any lean initiative in a higher education setting, one must evaluate and be cognizant of the cultural differences that have made lean conducive to the automobile industry as well as to manufacturing in general. Recognizing these differences at the start of any lean initiative may be a determining factor in whether or not a particular lean initiative succeeds in higher education. This research provides the basis for examining the cultural differences between the two sectors as they relate to lean initiatives.

10. Future research opportunities

In writing this article, a typical literature search did not result in a significant number of applicable articles. Therefore, a systematic literature review to determine the magnitude of peer-reviewed articles relating to cultural differences between higher education and the automotive industry would be beneficial to expand the knowledge. Additionally, the collection of data to further support the tables developed in this paper can substantially further the knowledge of cultural differences and the potential to apply lean to higher education.

11. References

Balzer, W.K. (2015). Where is the Evidence-Based Support for Lean Higher Education? The Research, the Challenge, and the Opportunities. *Conference presentation at Lean HE Hub Conference, University of Waterloo, Ontario*.

Balzer, W. K., Francis, D. E., Krehbiel, T. C., & Shea, N. (2016). A review and perspective on Lean in higher education. *Quality Assurance in Education*, 24(4), 442-462. DOI: 10.1108/QAE-03-2015-0011

Balzer, W.K. (2019). The role of Psychology in the development, application, and assessment of lean in higher education. *Conference presentation at Lean Higher Education 2019, University of Michigan, Ann Arbor.*

Balzer, W.K (2020). *Lean Higher Education*, 2nd edition. Routledge, New York, NY. https://doi.org/10.4324/9781351216944

Cano, M., O'Reilly, E., and Kourouklis, A. (2014). Lean second time around: Lessons learned for higher education. *17th Toulon-Verona International Conference on Excellence in Services. Liverpool, England, August 28-29*. http://sites.les.univr.it/eisic/wp-content/uploads/2018/07/Cano-O%E2%80%99Reilly-Kourouklis.pdf. September 29, 2019.

Emiliani, M. L. (2015). Engaging faculty in lean teaching. *International Journal of Lean Six Sigma*, 6(1). https://doi.org/10.1108/IJLSS-06-2014-0015.

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education*, 16(1), 3–36. doi:10.1080/13538321003679457.

Liker, J. (2021). *The Toyota Way; 14 Management principles from the world's greatest manufacturer* (2nd ed.). McGraw-Hill Company Inc.

Ohno, T., (2013). *Taiichi Ohno's Workplace Management: Special 100th birthday edition*. McGraw-Hill Company Inc.

Pascal, Dennis (2015). *Lean production simplified* (3rd ed.). CRC Press, Taylor and Francis Group, Boca Raton, FL.

Richter, L. (2011). Cargo Cult Lean. *Human Resources Management & Ergonomics,* 5(2), 84-93

Schein, E., & Schein, P. (2016). *Organizational culture and leadership* (5th ed.). The Jossey-Bass Business & Management Series.

Scoggin, S.C. (2017). The interdependency of Lean implementation and organization development. Dissertation. *The Chicago School of Professional Psychology*. Proquest Number: 10263384.

Shannon, C. (2020). Lessons from Implementing Lean at the Veterinary Teaching Hospital. In S. Yorkstone (Ed.), *Global lean for higher education: A themed anthology of case studies, approaches, and tools,* 311-328. Boca Raton, Florida: Taylor and Francis Group, LLC. https://doi.org/10.4324/9780429399688

Shannon, C. (2021). *Collaborative Problem Solving: A Guide to Improving Your Workplace*. Routledge, New York NY. https://doi.org/10.4324/9781003095057

Wiegel, V. (2020). *Lean in the classroom: The powerful strategy for improving student performance and developing efficient processes*. Routledge Productivity Press, New York NY. https://doi.org/10.4324/9780429837012

Touch Detection in Augmented Omni-Surface for Human-Robot Teaming

Fujian Yan[1]
Edgar Chavez[1]
Yimesker Yihun[2]
Hongsheng He[1*]
[1]*School of Computing, Wichita State University, USA*
[2]*Department of Mechanical Engineering, Wichita State University, USA*
Hongsheng.he@wichita.edu

Abstract

This paper proposes an architecture that augments arbitrary surfaces into an interactive touching interface. The proposed architecture can detect the number of touching fingertips of human operators by detecting and recognizing the fingertips with a convolutional neural network (CNN). The inputs of the CNN model are images that are ac[1]quired by an RGB-D sensor. The aligned depth information acquired by the RGB-D sensor generates the plane model, which determines whether fingertips are touching the surface or not. Instead of the traditional plane modeling method that can only be used for flat surfaces, the proposed system can also work on curved surfaces. Corresponding gestures are defined based on the detected touching fingers on the surface. The feedback of the robots is projected on the working surface accordingly with an interactive projector. Compared to conventional programming interfaces, directly touching is much more natural for human beings. Based on the experiments, the proposed system could reduce the massive training time of operators.

Keywords: Touching tracking, human-robot teaming, sensor fusion

1. Introduction

When robots operate in proximity to human workers, the relevance of human-machine, or more particularly, the human-robot interface grows. Compared to conventional industrial robots, which are used to working in a solo-working manner or static environments, advanced industrial robots are expected to collaborate with human co-workers. The study (Lasota & Shah, 2015) has indicated that the manufacturing process becomes quicker, more efficient, and less costly with human-robot collaboration. For most industrial robotic applications, a teaching pendant is provided to human co-workers for interacting with robots. Gestures are a common method of communication among human workers. They can also be used to interact with robots (Sheikholeslami, Moon, & Croft, 2017; Yan, Wang, & He, 2021).

For gestured-based human-robot collaboration interfaces, gesture ambiguity increases as the tasks get complicated. In contrast, touch interaction is unambiguous, tactile, familiar to users, and comfortable for extended periods (Li, Tan, & He, 2020; Li, Zhang, Li, & He, 2021; Xiao, Schwarz, Throm, Wilson, & Benko, 2018). New interaction models have emerged due to the advancements in display and vision technology, allowing for informative and real-time communication in shared workspaces. With the development of commercial RGB-D sensors, the prospect of converting these ordinary surfaces into

This work is supported by NSF CMMI 2129113.

Submitted June 30, 2022
Revised August 31, 2022

large, touchable, and interactive devices is increasing (Ntelidakis, Zabulis, Grammenos, & Koutlemanis, 2015; Yan, Tran, & He, 2020; Yan, Wang, & He, 2020). Therefore, a human-robot teaming method based on touching can benefit human-robot collaboration. For this purpose, this paper proposes an augmented Omni surface that can directly interact with robots.

To enable surfaces with interactive touching feedback, previous work has focused on refurbishing existing surfaces by adding other sensors such as acoustic sensors (Harrison & Hudson, 2008). Environments limit these reconstructed or refurbished surfaces. Two difficulties prevent human-robot teaming with a touching method without refurbishing the current environments, such as adding sensors to a current construction. The first problem is touch detection, and the second one is surfacing fitting. Touch detection includes two parts, fingertips detection, and contact detection. A conventional image processing method was proposed by (Bhuyan, Neog, & Kar, 2012) to detect and recognize fingertips. However, this method is limited by the complexity of the background (Xiao, Hudson, & Harrison, 2016). A deep learning approach is proposed by (Koller, Ney, & Bowden, 2016), where the number of visible fingertips in an image is random in particular cases. To detect each fingertip, different neural networks are trained (Alam, Islam, & Rahman, 2022). A touch detection method equipped with depth cameras is proposed by (Xiao et al., 2016), but this method is only suitable for flat surfaces. In particular, the working place for human-robot teaming has not only flat surfaces but also curved surfaces such as the carbine of a plane, the frame of a car, or the blades of a turbine.

In this paper, we propose a method that enhances human-robot teaming by directly interacting with robots on the surface of the working space. To adapt to different working environments, the touch detection method can enable humans to interact with robots by touch in different human-robot teaming environments, such as flat, curvy, convex, and concave surfaces with a single commercial RGB-D camera. Fingertips are detected with a convolutional neural network (CNN), and touch is detected based on the RGB-D sensor.

2. Augmented Omni-Surface

The developed Omni-surface architecture for human-robot teaming is shown in Figure 1. To enable augmented Omni-surface for human-robot teaming, touch detection is essential. The proposed augmented Omni surface superimposes mutual understanding between humans and robots on the surface of the working space.

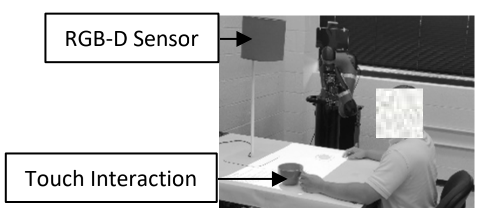


Figure 1. Augmented Omni-Surface.

The workflow of touch detection is illustrated in Figure 2. The proposed method contains two parts: surface fitting and touch detection. The RGB-D sensor captures an image having the user's hand, which will be the input for the fingertip's detection and surface modeling modules.

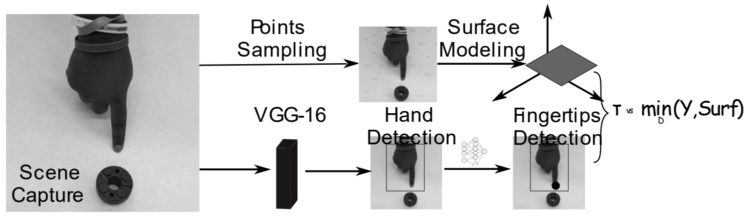


Figure 2. The framework of touch detection.

2.1. Arbitrary Surface Fitting

To fit the surface based on the collected depth scans, we assume the targeted surface is visible to the RGB-D camera. Once the camera has a clear sight of the plane, we collect random pixels $I(u, w, d)$ from the targeted surface. To fit the surface in a 3D space, we need to transform the 2D pixel coordinates into 3D world coordinates by

$$S(x, y, z)^T = K[R|t]^{-1} I(u, w, d)^T \tag{1}$$

where K is the intrinsic matrix of the RGB-D sensor and $[R|t]^{-1}$ is the inverse transformation matrix. With these random pixel coordinates, we can perform deprojection to calculate the depth of our random pixel coordinates. Specifically, our proposed method used Global Approximation (Piegl, 1996). In global surface approximation, surface error E and data are used as inputs to fit the approximation function. In general, the number of control points is not known to achieve the desired accuracy for the surface approximation. Thus, the approximation method is performed iteratively. In our proposed method, we used global surface approximation over the interpolation method due to the limited number of control points being automatically determined. Since surfaces are defined on a 2D plane, they can be described by

$$S'(\alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} N_{i,p}(\alpha) M_{j,q}(\beta) C_{i,j} \tag{2}$$

where α and β are parameters for the basis function $N_{i,p}$ and $N_{j,q}$ respectively. To determine a point using the α and β parameters, the basis functions for each of the parameters, $N_{i,p}(\alpha), and\ M_{j,q}(\beta)$, will be evaluated and multiplied with all the control points, $C_{i,j}$. In this sense, $N_{i,p}$ and $M_{j,q}$ are the knot vectors containing the size and degrees for our parameters.

The least-squares algorithm uses interpolation on the corner control points of the targeted surface to approximate the remaining control points of the approximated surface. All the parameters are used to create an approximation of the targeted surface. With the approximated surface, we begin to extract any curves that may exist in the approximated surface. This will return any curves that lie on the control points' surface.

2.2. Surface Touch Detection

Without installing additional hardware such as touch screens (Kim, Son, Lee, Kim, & Lee, 2013), it is challenging for the system to get feedback from human users' input. From an economic perspective and user-friendly, the proposed system is designed to detect touch by an RGB-D sensor. Typically, human beings operate touchable devices with fingertips, and fingertips have also played an essential role in human-robot interaction (Mitra & Acharya, 2007).

To detect fingertips from input images, two problems need to be solved. First, the hands need to be recognized, and second, we need to be able to retrieve the coordinates of each fingertip of the hand.

Earlier methods (Lai et al., 2016) isolate the hand regions from the acquired images using manually selected features such as color, depth, and contours. Some methods are not general, and the performance of these methods is easily affected by other natural factors such as the angles of taking images and ambiance lights (Nguyen, Kim, Kim, & Na, 2017). To obtain the coordinate of fingertips, the previous method (Liao, Zhou, Zhou, & Liang, 2012) employed a convex model that can model the endpoints of the extracted hand region. However, this method does not always find optimal results and is computationally costly (Jang, Noh, Chang, Kim, & Woo, 2015).

In this work, we formulate the problem of recognizing the hand and detecting the fingertips as a unified problem that combines classification and regression. We used a deep learning model to achieve this. To detect fingertips, we first detect the hand with 16 layers of visual geometry group (VGG) structures (Simonyan & Zisserman, 2014). Then, we recognize fingertips with a deep learning model that includes convolution layers, flatten layers, and coordinate layers. The structure of the fingertip detection model is shown in Figure 3. The model contains hand detection and fingertip recognition. Hands are detected by VGG-16 layers, and the output of VGG-16 is $\hat{Y}_h$. To extract features from $\hat{Y}_h$, a convolutional layer with the dimension of $(4 \times 4 \times 512)$ is employed. After the three convolutional layers, the output is $\hat{Y}_t$, which is a 2D array. Furthermore, it represents the positional information of the fingertips in the input image. To recognize fingertips, $\hat{Y}_t$ is flattened and fed into two fully connected (FC) layers.

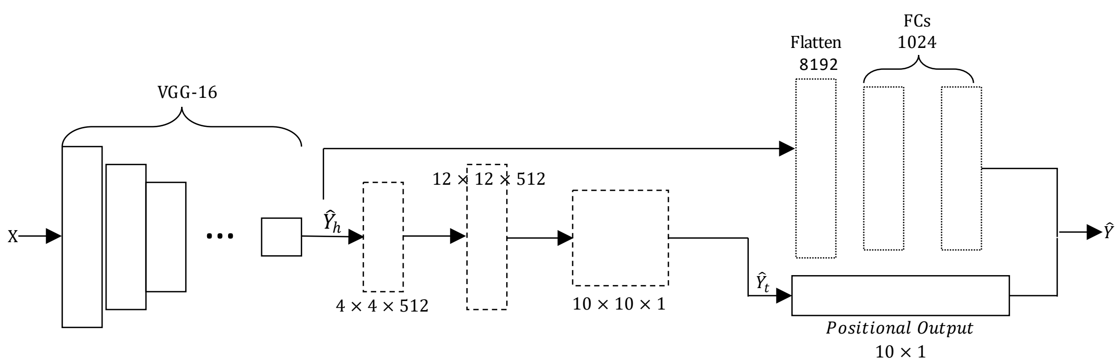


Figure 3. The structure of the model.

The input of the fingertip detection model is an image containing a single hand, and the model's output is the recognized fingers and their respective coordinates. The observed fingertips in the image are labeled as "1", and the unobserved fingertips are labeled as "0". The input of the model is $\mathrm{X}(u_i,\ w_i, f)$, where f is the label of each fingertip, and $f \in \{f_t, f_i, f_m, f_r, f_p\}$. The output of the model is $Y(u_i, w_i, P)$, where P is the probability of each detected fingertip's category, and $P \in \{P_t, P_i, P_m, P_r, P_p\}$. To minimize loss, the cross-entropy (CE) function and mean squared error (MSE) are employed with the loss

$$m = min\left[\boldsymbol{CE}\left(\widehat{\boldsymbol{P}}, \boldsymbol{P}\right) + \boldsymbol{MSE}\left(\widehat{\boldsymbol{Y}}(u_i, w_i)\ \boldsymbol{wi}, \boldsymbol{Y}(u_i, w_i)\right)\right] \tag{3}$$

where $\hat{P}$ and $\hat{Y}(\cdot)$ are the predicted results from the model, and P and $\mathrm{Y}(\cdot)$ is the coordinate of fingertips of labeled data used as ground truth.

To detect whether the user touched the surface, we computed the Euclidean distance between the detected fingertip $\hat{Y}(u_i,\ w_i)$ and the approximated surface $S_i(x, y, z),\ i \in \{0,1,2,3 \dots n\}$. We convert the detected fingertips in the image coordinates to the world frame coordinate $\hat{Y}(x_i, y_i, z_i)$, and then we calculate the distances between the detected fingertips and the approximated surface points, d_m, where $m = 1,2,3, \dots, m$. We found the minimum distance $d^*$and compared it with z_i. If $d^* \leq \tau +$

z_i, then the detected fingertips are touching the surface, and τ is the thickness of a human's fingers.

3. Experiment

To evaluate the performance of the method to augment Omni surfaces, we executed the proposed system in different scenarios to investigate the generality of the method. To investigate the performance, we have computed the accuracy of our touch detection method. We have also evaluated the user's satisfaction level based on the designed questionnaire.

3.1. Performance of Finger Touching

We tested the augmented Omni-surfaces method on different scenarios, and some of the results are shown in Figure 4. In this experiment, the augmented Omni-surfaces method has been applied to different scenarios, including flat wooden surfaces, laminated flat surfaces, cloth concave surfaces, convex Styrofoam surfaces, cardboard flat surfaces, curved plastic surfaces, and bumpy brick surfaces. These scenarios cover different shapes and materials of the surface. Our results conclude that the convex Styrofoam surface and the small cardboard had a false detection, while the rest of the other surfaces showed correct detection results. Therefore, the augmented Omni-surfaces method is reliable in practical usage. There are two fault-detected results that are grouped by the dashed rectangles in Figure 4. The reason for these fault results is that the working surface is small, and the depth collections include more noise than desired points. Therefore, the modeled surfaces are not correct. One solution would be to decrease the area of the collected points and increase the number of points collected.

To evaluate the performance of the proposed method, we have run simulations on different types of surfaces. For each type of surface, we simulated 10,000 points, including touching or not touching. We computed precision, recall, and F-1 scores. The results are shown in performance Table1. Compared with other types of surfaces, the curved surface has the lowest result in precision, recall, and F-1 scores. The reason for this is that the curved surface used in this experiment takes an extreme case to test the performance of the proposed model.

Table 1. Touch detection performance

Surface Type	Precision	Recall	F-1
Concave Surface	84.46%	100%	91.66%
Curved Surface	62.2%	97.24%	75.87%
Flat Surface	68.68%	98.16%	80.81%
Overall	71.78%	98.47%	82.78%

3.2. User Acceptance

To evaluate the acceptance of augmented Omni-surface interaction, we designed a questionnaire consisting of four questions with four evaluation criteria. These criteria are correctness, ambiguity, effectiveness, and complexity. We also assigned the weight for each criterion to compute the final score for each perspective. 1) The correctness is defined as whether the user can instruct the robot with the proposed interaction method; 2) The ambiguity is defined as whether the user can precisely give the robot's instructions; 3) Compared with other interaction methods such as programming, or physical interaction with buttons, the designed interaction is effective; 4) Compared with other interaction

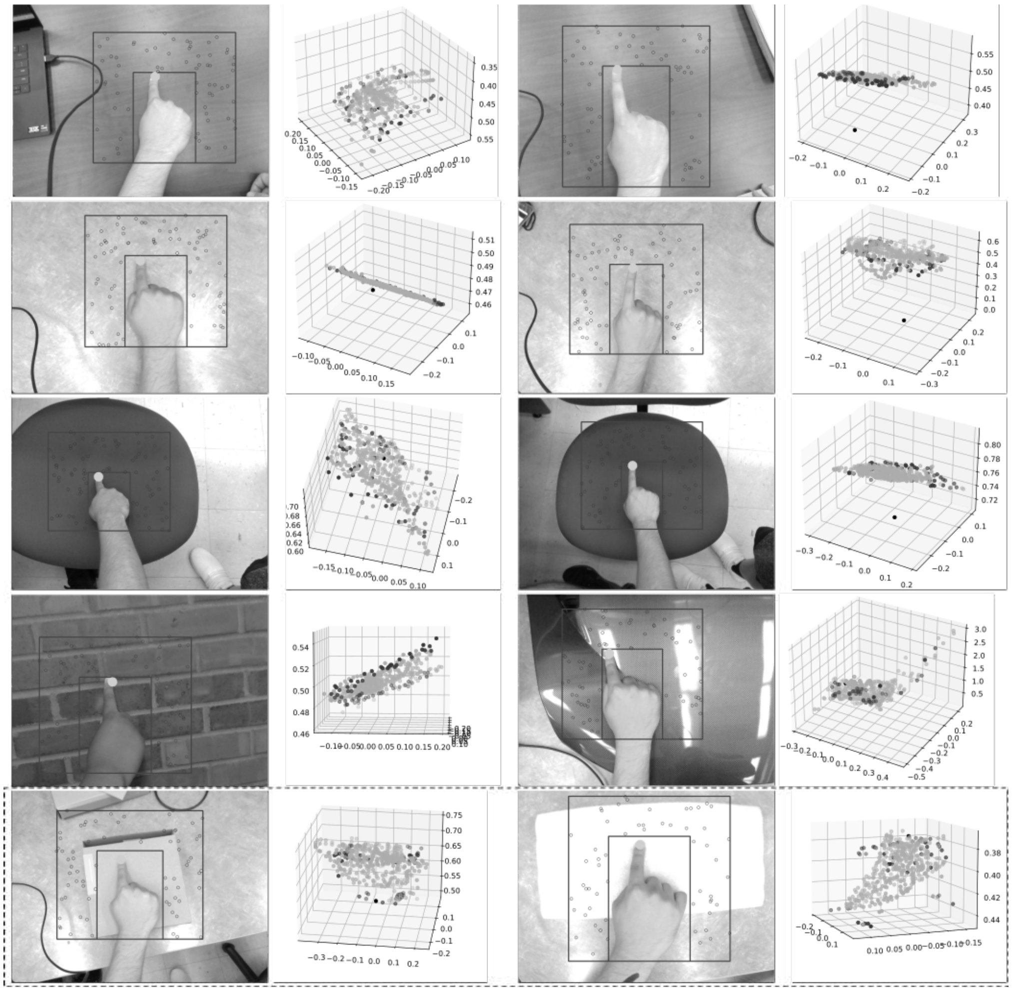


Figure 4. Touch detection results on different surfaces. The left image shows the detected results, and the right image shows the modeled surfaces. The bottom row demonstrated unsuccessful detections.

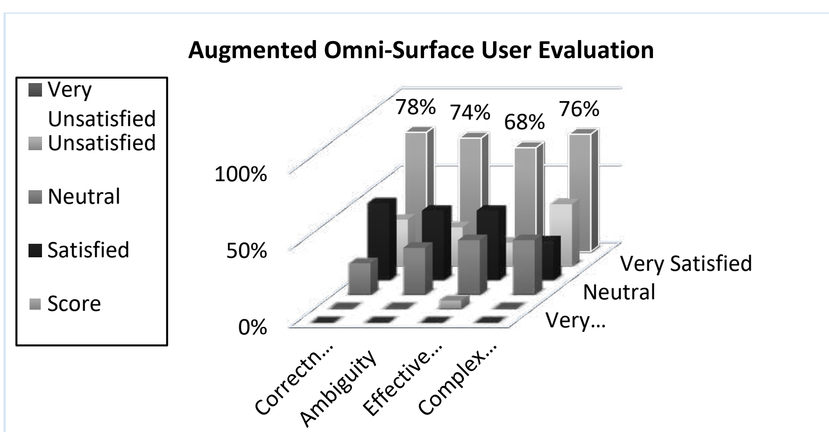


Figure 5. User satisfaction evaluation. There are five satisfaction levels for four evaluation criteria.

methods such as programming, or physical interaction with buttons, the designed interaction is complicated. Twenty participants took this questionnaire, and the results are shown in Figure 5. Based on the feedback from the questionnaires, users are satisfied with the proposed interaction method.

4. Conclusion

A method of touch detection for human-robot teaming has been proposed based on RGB-D sensing to detect a touch on arbitrary surfaces to enhance human-robot teaming. From the results of simulations, the proposed method achieved F-1 scores at 82.87%, which indicated the proposed method is promising. The proposed interface demonstrated improved user acceptance of traditional interaction methods according to the results of user satisfaction questionnaires.

5. Reference

Alam, M. M., Islam, M. T., & Rahman, S. M. M. (2022). Unified learning approach for egocentric hand gesture recognition and fingertip detection. In Pattern Recognition. 121, 108200. Elsevier BV. https://doi.org/10.1016/j.patcog.2021.108200

Bhuyan, M., Neog, D. R., & Kar, M. K. (2012). Fingertip detection for hand pose recognition. *International Journal on Computer Science and Engineering, 4*(3), 501.

Harrison, C., & Hudson, S. E. (2008). *Scratch Input: Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces.* Paper presented at the Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, New York, NY, USA. ACM Press. https://doi.org/10.1145/1449715.1449747

Jang, Y., Noh, S.-T., Chang, H. J., Kim, T.-K., & Woo, W. (2015). 3d finger cape: Clicking action and position estimation under self-occlusions in egocentric viewpoint. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 501-510. https://doi.org/10.1109/tvcg.2015.2391860

Kim, S., Son, J., Lee, G., Kim, H., & Lee, W. (2013). *TapBoard: making a touch screen keyboard more touchable.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 553-562. http://doi.org/10.1145/2470654.2470733

Koller, O., Ney, H., & Bowden, R. (2016). *Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA.

Lai, Z., Yao, Z., Wang, C., Liang, H., Chen, H., & Xia, W. (2016). *Fingertips detection and hand gesture recognition based on discrete curve evolution with a kinect sensor.* Paper presented at the 2016 Visual Communications and Image Processing (VCIP). Chengdu, China. 10.1109/VCIP.2016.7805464

Lasota, P. A., & Shah, J. A. (2015). Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration. *Human factors, 57*(1), 21-33. https://doi.org/10.1177/0018720814565188

Li, H., Tan, J., & He, H. (2020). *Magichand: Context-aware dexterous grasping using an anthropomorphic robotic hand.* Paper presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France. 10.1109/ICRA40945.2020.9196538

Li, H., Zhang, Y., Li, Y., & He, H. (2021). *Learning Task-Oriented Dexterous Grasping from Human Knowledge.* Paper presented at the 2021 IEEE International Conference on Robotics and Automation (ICRA).Xi'an, China. DOI: 10.1109/ICRA48506.2021.9562073

Liao, Y., Zhou, Y., Zhou, H., & Liang, Z. (2012, August 27-29). *Fingertips detection algorithm based on skin colour filtering and distance transformation.* Paper presented at the 2012 12th International Conference on Quality Software. Xi'an, China. https://doi.org/10.1109/QSIC.2012.62

Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37*(3), 311-324.

Nguyen, H. D., Kim, Y. C., Kim, S. H., & Na, I. S. (2017- July 29-31). *A method for fingertips detection using RGB-D image and convolution neural network.* Paper presented at the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Guilin, China. https://doi.org/10.1109/FSKD.2017.8393373

Ntelidakis, A., Zabulis, X., Grammenos, D., & Koutlemanis, P. (2015). *Lateral touch detection and localization for interactive, augmented planar surfaces.* Paper presented at the International symposium on visual computing. https://doi.org/10.1007/978-3-319-27857-5_50

Piegl, L. a. T., Wayne. (1996). *The NURBS book*. (1st ed.). Monographs in Visual Communication. Springer Berlin, Heidelberg

Sheikholeslami, S., Moon, A., & Croft, E. A. (2017). Cooperative gestures for industry: Exploring the efficacy of robot hand configurations in expression of instructional gestures for human-robot interaction. *The International Journal of Robotics Research, 36*(5-7), 699-720. https://doi.org/10.1177/0278364917709941

Simonyan, K., & Zisserman, A. J. a. p. a. (2014). Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556

Xiao, R., Hudson, S., & Harrison, C. (2016). *Direct: Making touch tracking on ordinary surfaces practical with hybrid depth-infrared sensing.* Paper presented at the Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces. https://doi.org/10.1145/2992154.2992173

Xiao, R., Schwarz, J., Throm, N., Wilson, A. D., & Benko, H. (2018). MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics, 24*(4), 1653-1660. 10.1109/TVCG.2018.2794222

Yan, F., Tran, D. M., & He, H. (2020). Robotic understanding of object semantics by referring to a dictionary. *International Journal of Social Robotics, 12*(6), 1251-1263. https://doi.org/10.1007/s12369-020-00657-6

Yan, F., Wang, D., & He, H. (2020). *Robotic understanding of spatial relationships using neural-logic learning.* Paper presented at the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, NV, USA. 10.1109/IROS45743.2020.9340917

Yan, F., Wang, D., & He, H. (2021). *Comprehension of Spatial Constraints by Neural Logic Learning from a Single RGB-D Scan.* Paper presented at the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic. 10.1109/IROS51168.2021.9635939

Statistical Process Control Implementations Within the Aviation Industry

Roger Merriman II
Department of Industrial, Systems, and Manufacturing Engineering,
Wichita State University, USA
Roger.merriman@wichita.edu

Abstract

In 1931, Walter Shewhart introduced his work in statistical process control (SPC) through his book, Economic Control of Quality of Manufactured Product. A common confusion in the use of SPC is that it does not fit every business or industry. Since Shewhart's work was published, various companies and industries have implemented such practices as a method to control their manufacturing quality of products, service quality, and design quality.

In 2017, a supply chain quality organization from a well-known business jet manufacturer discovered that vendors of its supply base were posting ostensibly statistical process control charts on metric walls in their facilities. The charts were not error-free. Out-of-date data placed on the charts, the preponderance appeared to be weeks or months old, and incorrect use of control and specification limits indicated there might be a gap. The purpose of this research was to discover why businesses in the aviation industry are not taking advantage of SPC practices and methodologies.

To identify the potential barriers inhibiting the deployment or growth of SPC implementations in the aviation industry, data was gathered from a survey of original equipment manufacturers (OEM) and members of the aviation supply base. The major challenges or barriers to implementing SPC in the aviation supply base were recognized as leadership commitment, cultural mindset and behaviors, understanding the benefits of SPC, and a lack of knowledge.

Keywords: SPC, Statistical Process Control, Implementations, Aviation Industry

1. Introduction

When Dr. Walter Shewhart first developed Statistical Process Control (SPC) techniques, he intended to help improve the quality of Western Electric's telephone hardware. Through the proper use of statistical methods, SPC is a vital aspect in monitoring, managing, maintaining, and enhancing the performance of a process (either manufacturing or service)(Montgomery, 2012). Dr. Shewhart began discussing and documenting SPC techniques in the 1920s and formally in 1931 when Shewhart introduced his work in SPC through his book, Economic Control of Quality of Manufactured Product(Shewhart, 1931). Since the publication of Shewhart's work, companies and industries have used such approaches to improve the quality of their product manufacture, service, and design.

Based on physical evidence from on-site audits, informal interviews conducted with OEMs at Nadcap conference meetings during the 2019 year, (Nadcap is an industry program focused on providing industry requirements that can be used regardless of the customer), and informal discussions with aviation industry leaders, this author suspects that the aerospace industry has a lower SPC implementation rate than most other large-scale manufacturing type industries.

1.1. What is SPC

SPC is a set of statistical tools used to investigate product, service, and process variation. SPC can also

Submitted July 2, 2022
Revised August 19, 2022

be used to verify that products or services offered satisfy consumer expectations at a low cost.

Variation can be categorized as either a common cause, naturally occurring variation, or special cause, meaning there is an assignable cause for the occurrence. Special cause variation arises from incidental causes and is not continuously present in the process. Common variation, on the other hand, is caused by inherent or known sources in the process. Consider taking a flight from Dallas, Texas to Denver, Colorado to demonstrate this notion. The flight is scheduled to depart Dallas at 7:04 p.m., according to the airline. As a passenger, you should board the plane today according to the airline's instructions; once everyone is on board, the jet taxis and takes off on schedule at 7:04 p.m. The airline crew's availability, the working aircraft itself, favorable weather, and the airport tower upholding the airline's required schedule are some of the elements that influence the take-off time. These variables are referred to as common cause variables since they are constant and known or planned for each flight. You take the same trip from Dallas to Denver a week later, but it does not leave until 7:55 p.m., which is more than 50 minutes late. This aircraft departs late due to a strong thunderstorm that prevents flights from taking off or landing at the airport until the storm passes through the area. The thunderstorm is the special cause variable that is influencing the take-off time. It is not a normal condition expected for a given trip and therefore is considered a special case.

As mentioned previously, SPC is a set of tools. Key tools in SPC are control charts, continuous improvement, and design of experiments. However, it is often associated with or can be more easily comprehended through the concept of control charts.

1.2. Popularization of SPC

Sampling inspection techniques made SPC popular with the military after World War II and again in the 1980s. Following the war, the semiconductor and automobile sectors adopted SPC techniques as a formula for keeping competitive in the global market, owing to Japan's high-quality output of products(The History of Quality, 2021).

In the 1980s it was the quality revolution that brought SPC back to life. Total Quality Management (TQM) and Six Sigma methodologies reintroduced the use and need for SPC tools and techniques. Powell (1995), indicates that TQM most likely started in 1949 with the formation of the Union of Japanese Scientists and Engineers which was a commitment to improving productivity and improving quality of life following the war. While it was started in the late 1940s, it was not until the early 1980s that America took notice and began utilizing the tools and techniques associated with TQM (Powell, 1995). In the early to mid-1980s, it was Motorola that developed the Six Sigma techniques and began utilizing them to improve the quality of their product lines (Buetow, 1996).

1.3. SPC in the aviation industry

Within the boundaries of the Advanced Product Quality Planning (APQP) and Production Part Approval Process (PPAP) processes acquired from the automotive sector, the aviation industry has recently developed a renewed interest in statistical process control tools and methodology. APQP is a system, which includes SPC tools and methodologies, intended to improve the design and delivery of products based on customer requirements while eliminating the risk of product failure. The sub-process PPAP ensures that the design responsible manufacturer’s APQP supply base delivers the same reliable parts (Sherman, 2021). Both systems work together to facilitate the production of conforming parts (parts that have been proven to meet design specifications) for the customer.

As customer quality expectations continue to increase, Original Equipment Manufacturers (OEMs) that produce the final product need a method to ensure quality is controlled at all levels of the supply chain. In the aviation industry, the International Aerospace Quality Group (IAQG) developed its own version of the APQP and PPAP through a new standard guidance titled AS9145 – Requirements for Advanced Product Quality Planning and Production Part Approval Process. The largest OEMs for commercial and military aircraft are beginning to mandate these requirements to their supply chain, therefore requiring suppliers to implement SPC or statistical quality control methods to maintain and improve their products.

The question remains is if their supply base is ready for such a mandate. As part of this research, we seek to understand the extent to which SPC is applied in aviation organizations. SPC procedures are used by both APQP and PPAP to ensure that parts and assemblies are supplied at a high-quality level and can fulfill the demands of their manufacturing lines. The focus of this research will be on aviation suppliers who work with commercial and private business jet OEMs.

2. Research Objectives and Methodology

2.1. Research gap

SPC is recognized as a critical set of tools and procedures for analyzing process behavior and assisting executives and staff in manufacturing environments in making real-time decisions (Elg et al., 2008). According to a previous review of existing literature (Merriman II, 2018) and the author's industry experience, SPC implementations in the aviation industry are limited and/or are not performed correctly. Three questions arise as a result. To begin, how widely is SPC used in industry today? Second, how widespread is SPC use in the aviation industry? Third, what are the barriers/challenges that impede aviation OEM and the supply chain's use of SPC methods?

2.2. Research objectives

The primary goal of this study is to discover why businesses in the aviation industry are not taking advantage of SPC practices and methodologies.

2.3. Research methodology

Surveys are used by businesses and researchers across all industries to help answer specific issues where data cannot simply be acquired without visiting a facility or having a discussion with an individual or team. Demographic, categorical, and open-ended questions were created for this research and distributed to the participants. Two surveys were conducted: one at the industry level, which was discussed in a previous publication (Merriman II, 2018), and another within the aviation supply chain, which will be discussed in this research. The goal of the surveys was to gather data to validate and understand the current condition of SPC in the aviation industry. Secondly, the survey results should also provide insight into success factors and implementation frameworks for those companies that have or continue to drive SPC within their quality management systems.

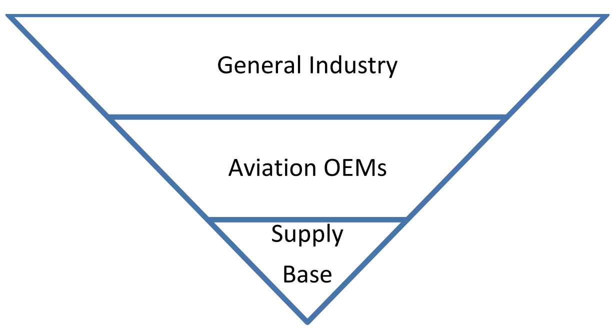


Figure 1 - Depiction of Survey Progression

Survey data was taken from both the OEM and the supply base within the aviation industry. OEM and supply base data were combined and used in the following analysis and discussion. Surveys were sent to 1475 participants using the survey software Qualtrics®. Two surveys requesting companies to disclose information and thoughts on the application of SPC were conducted. The first survey and review looked at large organizations that are engaged in quality practices through a large quality society, what was previously known as the American Society of Quality and is now just referred to as ASQ. This survey was used to develop additional questions for a second survey targeting input from OEMs and small to medium size organizations that are suppliers to larger aviation OEMs. Figure 1 is a visual representation of the survey scope, with the general industry survey being more widespread and less comprehensive than the aviation OEM and supply base surveys.

3. Survey Discussion

The data was broken down into six categories for the survey: demographics, potential barriers to SPC implementation, benefits of SPC, knowledge of SPC, training, and culture/leadership. This was developed utilizing information and implementations found in the literature.

3.1. Demographics

The survey invitation was sent to 22 OEMs and 1453 supplier-base organizations. 14 submissions were obtained out of 22 OEMs, while 305 responses were received out of 1453 suppliers. The aggregate response rate from OEMs and the supply base was 21.6 percent. According to the locations obtained from the survey system, the responses collected covered many areas inside the United States of America, areas of the United Kingdom, one response from Greece, and another from Mali, Africa. Figure 2 shows a graphical picture of the responses.

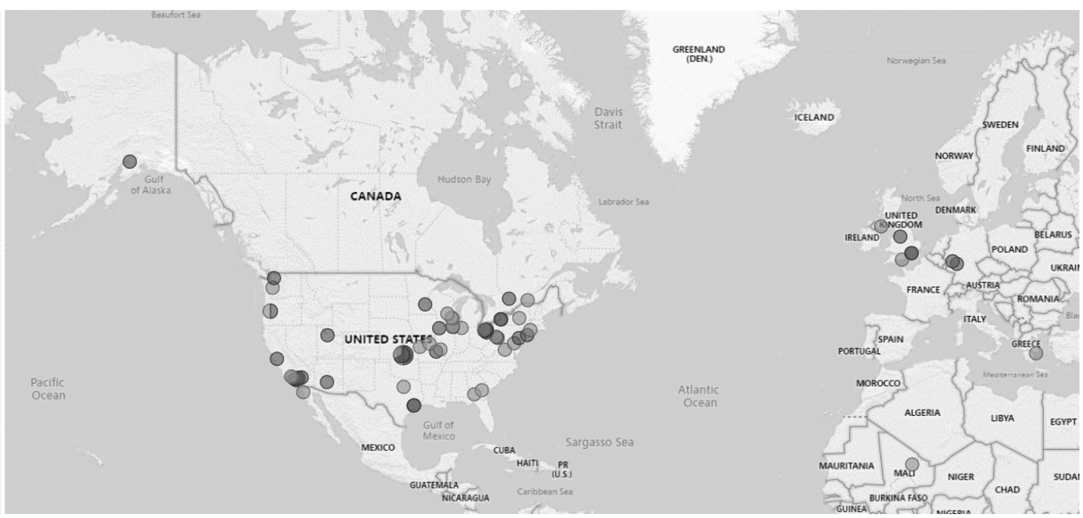


Figure 2 - Location map of aviation supply base survey participants

The response to whether SPC is employed in their operations is key to this study. 57.1 percent of OEMs said yes (shown in a light circle), while 42.9 percent said no (shown in a dark circle). 36.0 percent of suppliers said yes (light circle), while 64 percent said no (dark circle). Based on the observations and experience of this author, this result confirmed that there are more suppliers that do not employ SPC versus those that claim to have implemented SPC.

A common comment made concerning SPC is that "we can't implement because I don't have the resources available." OEM data were excluded from this data set because we are interested in the ability of the supplier to be able to implement SPC methods. The average number of employees within a given company from the responses was 173 (n=305), with a median of 90 and a range of 1497. Demographically this survey data represents companies with fewer than 1500 employees and may not characterize SPC applications in larger companies.

For those that said yes (n=110) to implementing SPC, they averaged 195 employees, a median of 115, and a range of 1497. For those that responded no (n=195), the average was 150 employees, a median of 100, and a range of 481. From the data provided by the survey responses and performing a t-test assuming unequal variances, the p-value is greater than 0.05 (p=0.259) which suggests that the averages are statistically the same and therefore there is no difference in the average number of employees that might suggest a reason that SPC could not be implemented. One might conclude that fewer employees might suggest there are not enough resources to support the implementation of SPC systems.

Reviewing the question related to revenue data, Figure 3, those with low revenue were still able to implement SPC and therefore revenue does not appear to be a barrier even though some businesses may claim this excuse. Those with lower revenue still found a way to implement SPC.

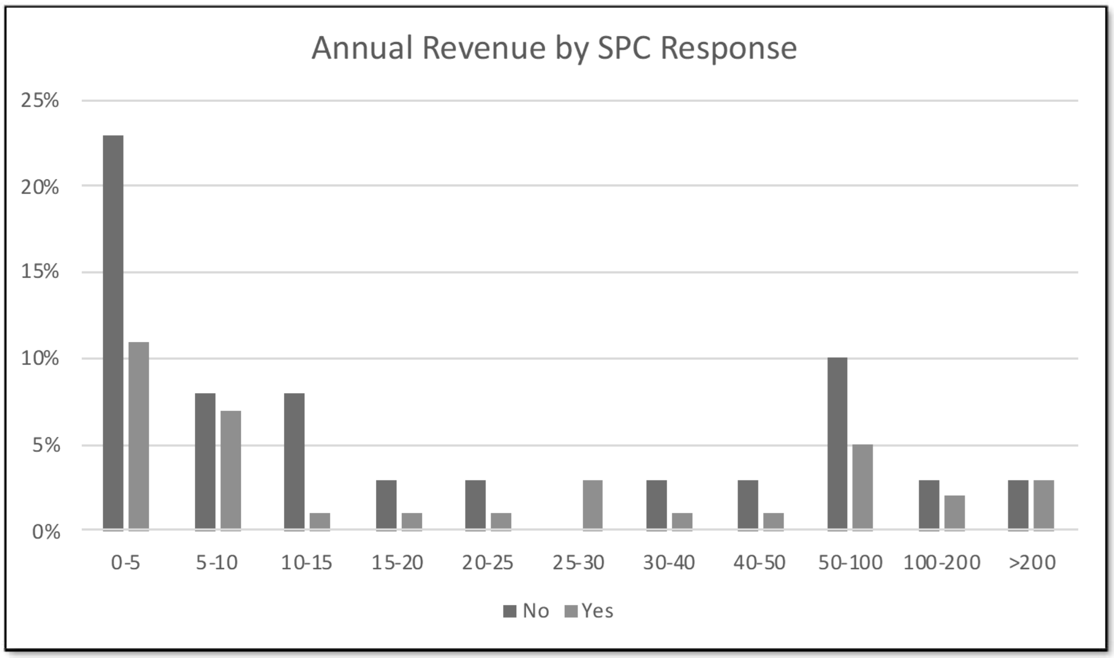


Figure 3 - Annual revenue ($M) by SPC response

Regarding those with quality system registrations, only 18% did not respond (see Table 1). 3% were from those that have implemented SPC and 15% from those that have not. With knowledge of the supply base that was invited to participate in the survey, it's plausible that the 15% include companies that have registration waivers. These businesses' OEMs may have determined that they are small enough or include another reason that they do not require registrations for their quality system. It's also conceivable that the individual who responded did not complete the question or was unaware of their specific registration(s). Any of the mentioned registrations do not specify a process control method, but they do request that the business be able to demonstrate that quality is regulated through a method of control. If they said “no to employing SPC methods”, on-site visits or interviews would be required to learn more about the context or methods they use to control their product quality.

Table 1 - Quality system registrations held

SPC Response	ISO 9001	AS9100	Nadcap	TS 16949	Other	No response
No	38%	39%	18%	1%	6%	15%
Yes	25%	29%	15%	2%	10%	3%

The last question with respect to demographics is interested in the position of the person who responded to the survey. For the responses that were received, 99 out of 105 have titles that are quality-related, of the remaining 6 respondents, 2 were from the finance organization, and 4 were from operations or engineering-type positions. The invites were sent to all the company's quality managers, directors, and presidents. It was not anticipated to be forwarded to the finance group for a response; yet, while 2 percent may not seem like a large number, it could add to data inaccuracy. Engineering and operations responsibilities may be incorporated in the quality role (or vice versa) as a second position for that employee, according to the author's experience working at a supplier. Employees who work in small businesses can and often do take on multiple responsibilities.

3.2. Barriers to implementing SPC

Impediments to implementation and barriers to sustaining or growing SPC inside their businesses were the subjects of two inquiries. One question was for companies that had not yet implemented SPC, and the other was for those who had. This would allow them to identify what was preventing them from expanding their implementation.

Lack of benefits to the business was ranked first, lack of knowledge was second, resource constraints were third, and training was fourth among the items highlighted as potential hurdles to implementing SPC. Leadership and culture were also hindrances to further initiatives. (See Table 2)

Table 2 - Potential barriers to implementation

Barrier Category	% Response
Lack of benefits	25%
No comment	21%
Lack of knowledge	14%
Resource constraints	9%
Attempting*	9%
Attempted	7%
Customer Requirements	5%
Other methods of control	5%
Process Constraints	2%
Business Type	2%

It was also intriguing to see that 9% were in the process of attempting to implement SPC methods. If the survey was given later, it would be useful to see if they were successful or if they are still postponing the effort.

3.3. Benefits of implementing SPC

The following section discusses the potential benefits of employing SPC procedures for organizations. We start by figuring out why they're employing SPC approaches in the first place. Continuous improvement, process performance, process control, and meeting customer or ISO 9000/AS9100 (International Organization for Standardization/Aerospace Standardization) requirements cover approximately 80% of the responses. If those who stated that the aim is to achieve ISO/AS criteria believe that the instruction requires SPC methods expressly for managing their product quality, they may not fully comprehend the quality system requirements. SPC is recommended by the standards, although businesses are free to pick their own method of product control.

Except for “process trial charts being utilized to analyze a process for a short duration,” the other responses are equivalent to findings in the literature. No additional clarification was received for the “other” response.

The top 80% of replies to the next question, which specifically asked about benefits, all relate to processing activities, ranging from performance measures (19.5%) to process enhancements (2.2%). All are valid responses and should be considered in communication plans related to educating the employee base with respect to the advantages of SPC methods.

3.4. Knowledge

The questions in the survey section related to the knowledge base, contain topics relating to SPC, and were designed to gauge supply base expertise.

This section covered different types of control charts. Shewhart charts, X-bar and R, P charts, C charts, U charts, and subtle variations of those charts are regarded as easier charts to apply to processes for this study. More complex techniques include CuSum, G Charts, Tool Wear, and EWMA. (See **Table *3***) Shewhart level charts are used to control processes in the aviation supply base, according to 62.7% of responses. This means that the implementations appear to be at a basic level. It does not indicate whether the implementations are correct or incorrect. Other findings could be that suppliers lack subject matter expertise to use more sophisticated charts, that processes do not require more advanced charts, or that the process requires an advanced chart, and the existing graphic is inaccurate.

Table 3 - Type of control charts used

Control Chart Used	Responses	Control Chart Used	Responses
X-bar and R	21	nP chart	2
Median and Range	14	C Chart	2
Individuals	10	P' Chart	2
Acceptance Chart / Acceptance Individuals	10	Tool wear / Tool wear Individuals	2
P chart	9	G chart	2
Moving Average / Moving Average Individuals	9	U Chart	1
X-bar and S	5	CuSum / CuSum Individuals	1
X-bar and S-squared	5	T squared	1
Other	5	T chart	1

Follow-on questions related to the development and operation of control charts were then asked of the participants. These responses indicate the level of expertise within the supply base which includes 8 responses from OEMs. Answers were combined to avoid singularities in responses at the OEM level.

Of interest was a response concerning how control charts are set up and maintained. One comment that was made in the "other" choice indicated that limits, sampling, and targets were adjusted on a year-by-year basis. Control charts limits and sampling plans are meant to be adjusted based on the process performance, not based on some timeframe changes unless there is a change in other process parameters or the product being measured has changed. Past performance is used to create the control limits and with a process starting over at the beginning of the year, this comment implies that they may be using limits other than traditional Shewhart calculated limits. This is an indicator that knowledge is suspect.

Another remark in the “other” response on control chart rules for out-of-control signals suggests that control limits are set using shop tolerances. While control limits can be established in a variety of ways, classic Shewhart charts only employ three standard deviations from the mean. The use of shop tolerances or specification limits is tied to the voice of the customer rather than the voice of the process.

Agreements or disagreements with the statements in question about the application of charts, which uses a 5-point Likert scale, indicate gaps in the knowledge base for implementations of SPC methods. Of these questions, updating charts directly ties in with the visual indications seen by this author while conducting audits within the supply base. Delayed, lack of, or no updating of charts is a good indicator that the charts may just be used to satisfy a customer's requirements. At the very least, they should be updating the data in preparation for the customers’ audit. They should reconsider using SPC methods or seek assistance in properly implementing them.

3.5. Training

Questions related to training in the survey were used to discover more about who is trained and who conducts the training within a business. Businesses frequently complain that they are resource constrained and thus unable to implement SPC. Not all staff are trained in SPC methods, according to 71.8 percent of the responses.

However, there was some encouragement that 52.6% of the respondents train employees on how to use control charts. This should indicate the availability of a local subject matter expert. It would be even better if the trainer's knowledge could be verified by a test or industry certification. Suppliers may lack the resources for this type of support, or they have found a way to accommodate this with their current employees since 7.7% of replies state they use a 3rd party or consultant for SPC training.

Following up with technical support, 64% say they can usually find technical assistance to resolve control charting concerns. It's unclear what kind of media or human assistance is being used to fulfill this requirement. Those implementing SPC should have access to a variety of resources. Different strategies can be used to solve various problems.

3.6. Culture and leadership

The last set of questions focuses on leadership and culture, two of the most significant aspects of a successful SPC implementation. Who is accountable for SPC, what measures are taken when processes get out of control, and what resources are available to remedy issues before resuming a process are all signs of whether leadership and the culture appreciate the need of controlling variance in their processes.

Ostensible practitioners of SPC claim that they just know what good data looks like. 18% of the responders claim this ability. Decisions should be based on data and specifically what is indicated by the use of a properly set up control chart. Knowing what good data looks like (as expressed in open-ended comments), implies that these users are going by intuition. Leadership should push back on these individuals and seek qualified support; otherwise, the employees' trust will be lost, which could result in poor implementation.

Support and commitment from the top are critical to the success of an SPC implementation. When an out-of-control condition arises, leadership that does not empower employees to halt the process from producing defects will result in the falsification of data on the charts or they will not be updated on the required frequency. According to the survey results, using questions related to leadership support, somewhat disagree and strongly disagree responses, 20.8% believe they are not empowered to take control of the process when the need arises. Using the same questions and *somewhat agree and neither agree nor disagree* responses, 54.6% only partially agree that they are properly supported by leadership. This poses a threat to the SPC system's long-term viability. The importance of leadership commitment in starting and maintaining the SPC implementation process has been explored in the literature (Merriman, 2018). If there is a problem with the implementation process, resources will most likely not be directed towards addressing the issue without commitment and the support of the business.

The last question relates to the culture of the business. Older quality systems rely on a typical quality department to facilitate and maintain quality initiatives. Newer systems are driving everyone to play a role in and be responsible for their own quality. Unfortunately, charting the data and evaluating the data generated higher quality responsible responses than other groups. Obtaining the data and acting on the data indicated responsibility lies with other functional areas. Acting on the data shows that management is becoming more involved in the response, although management was not defined explicitly, which could still indicate that quality management oversees corrective action. See Table 4.

Table 4 - Who is authorized to perform the following steps with control charts

	Collect the Data	Plot the Data	Analyze the Data	Act on the Data
Quality	34.3%	64.7%	45.7%	25.7%
Engineering	2.9%	5.9%	14.3%	11.4%
Operator	54.3%	17.7%	11.4%	14.3%
Management	8.6%	11.8%	28.6%	48.6%

4. Summary and Conclusions

The purpose of this research was to figure out why suppliers in the aviation OEM and supply chain do not use SPC techniques and methodologies as frequently as other industries. Interest in SPC methods for managing product quality has developed in numerous industries since the implementation of TQM and Six Sigma in the 1980s and 1990s, but it does not appear that aviation has joined the frenzy. Effective frameworks for implementing SPC based on success factors have been developed and assessed in industries such as food, health, and software, according to a review of the literature in a previous study (Merriman II, 2018). A framework for the aircraft sector has not been devised, according to literature searches. Survey responses from various aviation suppliers and OEMs, as well as an analysis of frameworks and success factors from other industries, will be needed to complete the research to start developing a conceptual framework for implementing SPC methodologies within the aviation supplier community.

This research identified the following elements that emerged as barriers to implementing SPC in the aviation industry from taking advantage of SPC:

- To create a culture and environment conducive to SPC implementation in the supplier base, leadership commitment is insufficient.
- Within the aviation industry, the benefits of SPC are not generally known or understood, resulting in a gap.
- SPC adoption is not supported by cultural beliefs and practices. Leadership does not always create an environment in which employees are empowered to make the best decision possible based on process data.
- A lack of understanding of SPC is insufficient to justify the use of SPC tools and techniques.

The identified gaps and barriers will be employed to construct conceptual training and implementation frameworks to assist the aviation supply base with SPC implementations in the future (Merriman, 2021). Pilot studies with a sample of aviation suppliers should be used to verify these frameworks. To prepare for new deployments, the frameworks should be improved on a regular basis.

4.1. Future research

This study's research is only the beginning of the process. Working with suppliers to develop a framework, along with a training program was limited due to Covid constraints during a portion of this research and is intended for the future when restrictions are lifted. Implementations are necessary to validate the structure of frameworks. The preference of this author is to avoid confirming Woodall's (2000) remark on the disparity between theoretical development and implementations of SPC. The next steps in research should involve the implementation of an SPC system at selected suppliers to validate and refine the frameworks discussed. This will also allow for the development of process details needed for a successful implementation.

4.2. Limitations

Surveys are not without limitations. Other populations may be of interest in understanding why companies do not or struggle to implement SPC. This set of surveys also utilized open-ended questioning without limited follow-up or clarification of the responses submitted. Understanding context and situational awareness will give better insight into any gaps that may exist. Incomplete responses are limiting when the thought being conveyed is not clear and the analysis of the response could lead to an incorrect conclusion. Sample size and demographics can influence results from one spectrum to another and might lead to false or incomplete conclusions. While the surveys discussed in this research had a return rate of 21.6%, there is always the potential that this sample of data could still be driven in other directions along the spectrum of other implementations.

The COVID-19 pandemic prompted a decision to focus on survey responses and literature research, deferring a visit to the supply base to gain a better understanding of the needs of those businesses. It's also possible that not conducting follow-on visits or discussions may have harmed one's ability to learn more through face-to-face conversations. The risk is limited, however, because the survey results were frequently confirmed by existing literature.

References

Buetow, R. (1996). The Motorola quality process: Six sigma. *Cases_in_Quality. GM Bounds. Chicago, Irwin*, 229-249.

Elg, M., Olsson, J., & Dahlgaard, J. J. (2008). Implementing statistical process control: an organizational perspective. *International Journal of Quality & Reliability Management*.

Merriman II, R. A. (2018). A Review of Current Implementations of Statistical Process Control in Large Organizations. *Journal of Management & Engineering Integration*, *11*(2), 46-54.

Merriman, R. A. (2021). *Statistical process control implementations within the aviation industry* Wichita State University].

Montgomery, D. C. (2012). *Statistical quality control*. Wiley Global Education.

Powell, T. C. (1995). Total quality management as competitive advantage: a review and empirical study. *Strategic management journal*, *16*(1), 15-37.

Sherman, F. (2021, October 25, 2018). *Difference Between APQP and PPAP*. BizFluent. Retrieved June, 17, 2021 from https://bizfluent.com/info-8712244-difference-between-apqp-ppap.html

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Macmillan And Co Ltd.

The History of Quality. (2021). ASQ. Retrieved Feb. 2, 2021) from https://asq.org/quality-resources/history-of-quality

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, *32*(4), 341-350.

Computer Applications: A Novel Architecture to Improve the Performance of Audio Visual Applications

Abu Asaduzzaman[1*]
Kishore K. Chidella[2]
Fadi N. Sibai[3]
[1]*Wichita State University, USA*
[2]*University of Nevada at Las Vegas, USA*
[3]*American International University, Kuwait*
abu.asaduzzaman@wichita.edu

Abstract

To attain the best audio-visual experience, the underlying computing platform should provide high performance in real time. The forthcoming computer systems consist of multicore processors to achieve a high performance-to-power ratio. The algorithms in visual computing are becoming highly complex to meet the requirements. According to the new design paradigm, a simultaneous exploration of multicore architecture and complicated algorithms is beneficial to achieve the design goals for visual systems. However, caches in multicore architecture multiply the timing unpredictability and that creates a serious challenge in running real-time audio/visual applications in multicore systems. In this work, we introduce a novel multicore architecture with miss tables inside level-1 caches to improve performance and decrease power consumption. Miss table holds block address information regarding the application being processed that causes cache misses. Miss table information is used for efficient selection of the blocks to be locked or victim blocks to be replaced. This approach improves predictability by locking important blocks inside the cache during the execution time. At the same time, this approach decreases average delay per task and total power consumption by reducing cache misses when the right cache blocks are locked and/or replaced. We simulate an 8-core architecture that has 2 levels of caches using the Moving Picture Experts Group (MPEG)-4 decoding and Fast Fourier Transform (FFT) workloads. Simulation results show that a reduction of 42% in mean delay per task and a reduction of 40% in total power consumption are achieved by locking 20% of the total level-1 instruction (I1) cache size.

Keywords: Computer applications, multicore architecture, computing performance, cache memory hierarchy, audio-visual applications

1. Introduction

Visual computing uses a visual paradigm rather than a conventional text paradigm and deals with large image files including video sequences (Visual Computing, 2017). The algorithms required for future visual systems are much more complicated than those currently being used. Current single-core processors are not adequate to support future visual computing. According to the newly emerged

Submitted August 14, 2022
Revised September 7, 2022

multicore design, two cores running at one-half of the frequency can approach the performance of a single core running at full frequency, while the dual-core consumes significantly less amount of power. Because of their high performance/power ratio, the multicore processors open new possibilities for system designers in implementing highly complex visual computing algorithms (Suhendra and Mitra, 2008; Chiappetta, 2007). In a multicore architecture, two or more independent cores are combined into a die. In most cases, each processor has its own private level-1 cache memory (CL1). Normally, the CL1 is split into instruction (I1) and data (D1) caches. Also, multicore processors may have one shared level-2 cache (CL2) or multiple distributed and dedicated CL2s (Burt, 2009; Online: Wikipedia, 2022). Cache parameters (such as cache size, line size, and associativity levels) significantly influence system performance (Asaduzzaman et al., 2006). Multicore architectures are more suitable for real-time audio/visual applications because the concurrent execution of tasks on a single processor is inadequate for achieving the required level of performance and reliability. The integration of billions of transistors in a single chip is now possible. As a result, the multicore design trend is expected to grow for the next decade. However, the cache memory introduces timing unpredictability and real-time applications demand timing predictability and cannot afford to miss deadlines. Therefore, it becomes a great challenge to support real-time multimedia applications on multicore systems. Knowing that performance can be sacrificed in real-time systems, a better use of the cache can be very useful to improve predictability and performance and reduce total power consumption.

Cache locking shows promises to improve predictability (Asaduzzaman et al., 2007; Puaut, 2006; Tarui et al., 1992). Cache locking is defined as the ability to put off some or all of the data or instruction cache from being overwritten. Cache entries can be locked either for individual ways within the cache or for the entire cache. In way locking, only a portion of the cache is locked by locking ways within the cache. Unlocked ways of the cache behave normally. Way locking improves predictability and performance; using way locking, the Intel Xeon processor achieves the effect of using local storage by the Synergistic Processing Elements (SPEs) in the International Business Machines (IBM) Cell processor architecture (Thompson, 2012). In entire cache locking, cache hits are treated in the same manner as hits to an unlocked cache. Cache misses are treated as cache-inhibited access. Invalid cache entries at the time of locking will remain invalid and inaccessible until the cache is unlocked. Entire cache locking is inefficient when the data size or instructions to be locked is smaller compared to the cache size.

In currently available single-core way cache locking techniques, the blocks to be locked are selected by offline analysis of the applications. The information collected by offline analysis can be post-processed and re-used in single-core and/or multicore systems to facilitate the efficient selection of blocks to be locked and/or victim blocks to be replaced. In this work, we introduce miss tables inside level-1 caches of multicore processors to hold block address information that causes most or all cache misses. This approach improves predictability and decreases average delay per task and total power consumption by reducing cache misses as the right cache blocks are locked and/or replaced.

Paper Organization: In Section 2, related literature is discussed. The proposed multicore architecture with miss tables in CL1s is introduced in Section 3. In Section 4, the proposed scheme is evaluated by presenting important simulation results. Finally, this work is concluded in Section 5.

2. Literature Survey

Predictability and performance improvement for visual computing systems with cache memories has become an important research topic in the past years. A lot of work has been done to improve predictability in single-core systems by cache locking. In this section, we first briefly discuss cache memory hierarchy, followed by some existing single-core cache-locking techniques. Then, we present some cache memory hierarchies used by contemporary popular multicore processors. In the following section, we introduce our proposed multicore architecture with miss tables in CL1s.

Cache memory has a very rich history in the evolution of modern computing (Lutkevich, 2020). Cache memory is first seen in the IBM System/360 Model 85 in the late 1960s. In 1989, Intel 468DX microprocessor introduced an on-chip 8 KB CL1 cache for the first time. In the early 1990s, off-chip CL2 cache appeared with 486DX4 and Pentium microprocessor chips. Today's microprocessors usually have 128 KB or more of CL1, 512 KB or more of CL2, and an optional 2 MB or more CL3. Some CL1 cache is split into I1 and D1 to improve performance (Inkley et al., 2006). Intel Pentium 4 processor, one of the most popular single-core processors that use inclusive cache architecture, is discussed in (Pentium 4, 2022). A typical inclusive cache is shown in Figure 1. CL2 contains each and every block that CL1 (i.e., I1 and D1) may contain. In case of a CL1 miss followed by a CL2 miss, the block is first brought into CL2 from main memory, then into CL1 from CL2. Intel Pentium 4

Willamette is a single-core processor that has an on-die 256 KB inclusive level-2 cache; with an 8 KB level-1 trace/instruction cache (I1) and 8 KB level-1 data cache (D1).

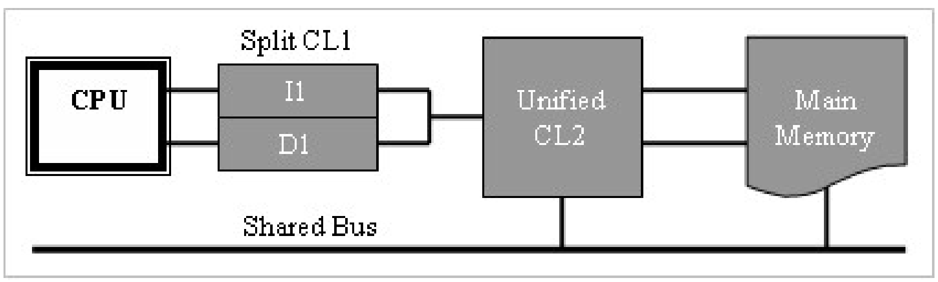


Figure 1: Inclusive cache architecture

In (Vera et al., 2003), a memory hierarchy with a dynamic locking cache is proposed to provide high performance combined with high predictability for complex systems. They conclude in an intuitive way that faster execution times are possible. However, it is very difficult to provide the necessary locking information into the program code since both the hardware and the software might be involved. In (Arnaud et al., 2005), a methodology using a genetic algorithm is proposed that can be used to select a set of instructions to be preloaded and locked in the cache. This scheme may improve predictability by selecting the target set of instructions efficiently. However, the algorithm used in this scheme can be also misleading, because it uses neither the information about task structure nor the problem parameters. In (Tamura et al., 2005), the impact of three different fitness functions is studied. These fitness functions are used in a genetic algorithm that selects the contents of a static locking cache memory in a real-time system. Results indicate that none of the fitness functions perform well in the two basic metrics used. In (Tamura et al., 2004), static cache analysis is combined with data cache locking to estimate the worst-case memory performance in a safe, tight, and fast way. Experimental results show that this scheme is more predictable. However, a better analysis that classifies the cache accesses as misses or hits and locks fewer regions could be beneficial. In (Campoy et al., 2005), an algorithm proposed dynamic cache locking which partitions the task into a set of regions. Each region owns statically a locked cache content determined offline. A sharp improvement is observed, as compared with a system without caches. However, this technique is not capable of power estimation – a crucial design factor for embedded systems. Techniques discussed here are developed to evaluate predictability in a single-core system and they are not adequate to analyze performance, power consumption, and predictability of multicore real-time systems.

Most manufacturers are adopting multicore processors to acquire the required high processing power for future visual computing systems. Popular multicore processors from Intel, AMD, and IBM have multilevel caches (Inkley et al., 2006; Chiappetta, 2007; Burt, 2009; Online: Wikipedia, 2022). AMD quad-core Opteron has 256 KB I1, 256 KB D1, 2 MB dedicated CL2, and 2 MB (Santa Rosa) or 4 MB (Deerhound) shared CL3. As shown in Figure 2, the Intel quad-core Xeon DP family has 128 KB I1, 128 KB D1, and 8 MB shared CL2 (Inkley et al., 2006).

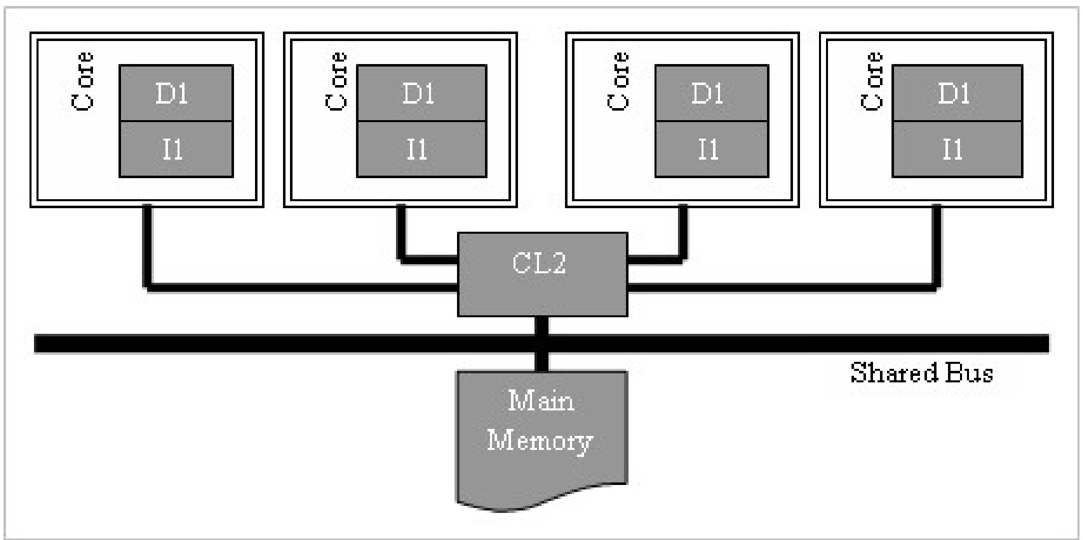


Figure 2: Intel quad-core Xeon DP family architecture

IBM's Cell multicore processor has a Primary Processing Entity (PPE) like IBM dual-threaded PowerPC and 8 Cells. Figure 3 shows a Cell-like multicore architecture (Blachford, 2005; Vance, 2006). Each Cell is also called a Synergistic Processing Element (SPE). The PPE contains a 32 KB I1 and a 32 KB D1 cache. A 512 KB CL2 is shared by the PPE and SPEs. Primarily the PowerPC PPE keeps the processor compatible with lots of applications. Each SPE has 256 KB static random access memory (SRAM) and a 4 x 128-bit ALU (Arithmetic Logical Unit which does the math in a processor), and 128 of 128-bit registers. The Element Interconnect Bus (EIB) is the communication bus internal to the Cell processor which connects the various on-chip system elements: PPE processor, memory controller (MIC), eight SPE coprocessors, and two off-chip I/O interfaces.

Various shared and distributed cache memory organizations and the impact of sharing and privatizing them on performance in homogeneous multicore architectures are studied in (Sibai, 2008).

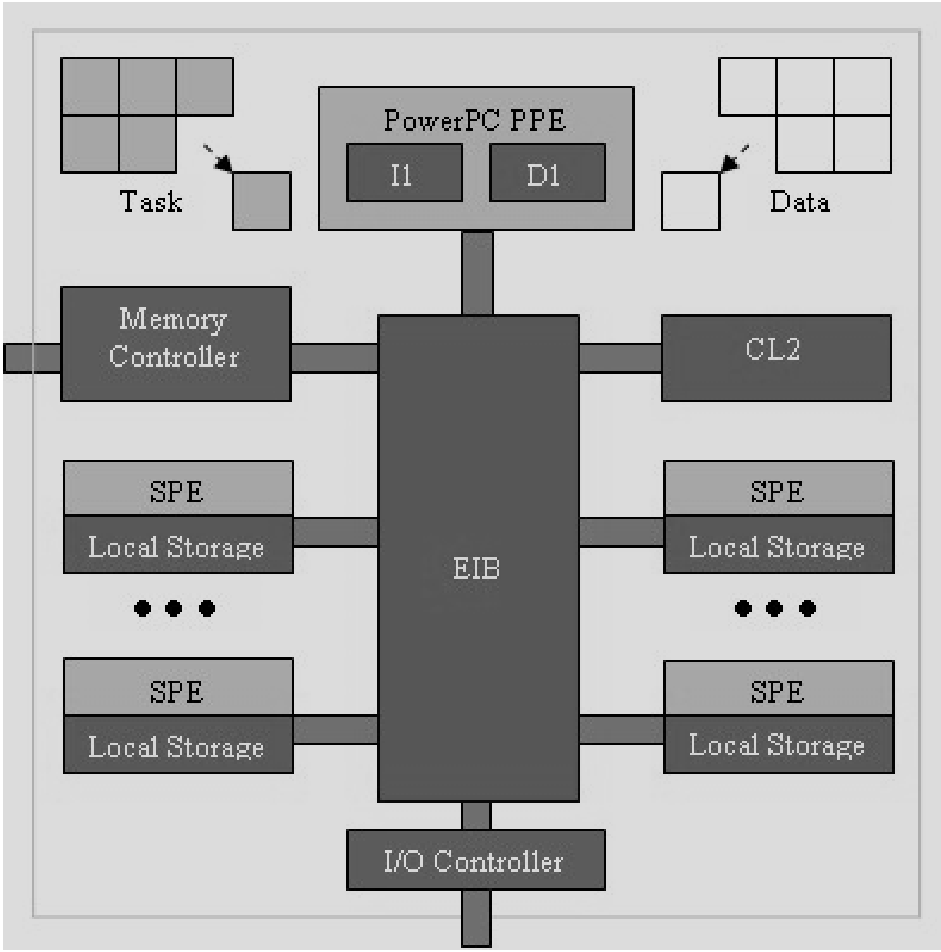


Figure 3: IBM Cell-like multicore architecture

3. Proposed Multicore Architecture with Miss Tables in CL1s

We have discussed the cache memory hierarchy in some popular single-core and multicore processors and some existing cache-locking techniques. In this section, we introduce our proposed multicore architecture with miss tables (MT) in CL1s. The miss table is populated with the block information that causes cache misses by post-processing the results of the worst-case execution time (WCET) analysis of the applications. Using MT, the proposed architecture can take advantage of multicore cache locking, selective pre-loading, and a better cache replacement strategy. In the following subsections, we discuss the architecture, miss table, cache replacement policy, and multicore cache locking algorithm.

3.1 Architecture

Figure 4 illustrates the proposed multicore architecture with 8 cores. Along with MT, each core has one private CL1 which is split into I1 and D1 for improved performance. We use Intel-like shared CL2 in our proposed architecture. CL2 may be partitioned into parts to reduce bus contention so that only a few cores can access each part.

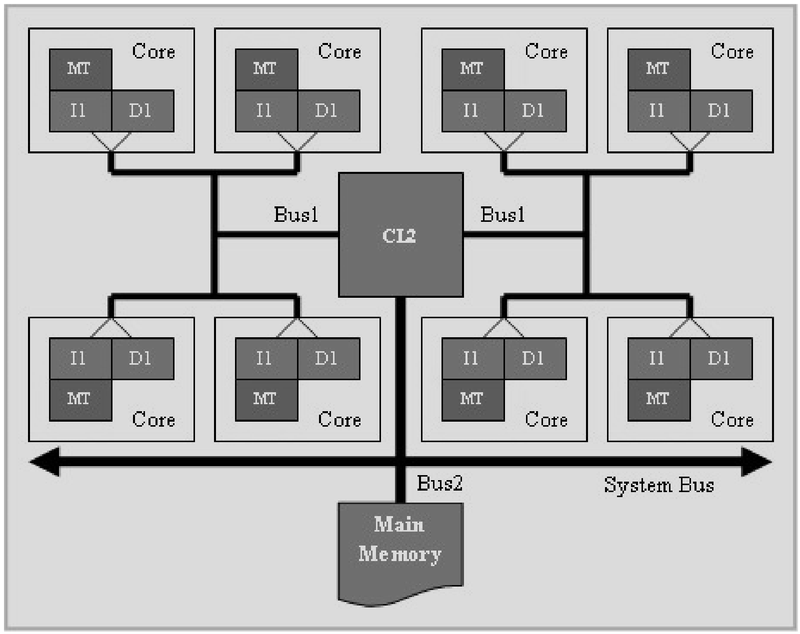


Figure 4: Simulated 8-core architecture with MT

3.2 Miss table

A miss table (MT) is a table that contains the addresses of all or top-most blocks that cause cache misses – block addresses are sorted in descending order of the miss numbers. For each application, after post-processing the tree graph (generated by Heptane), one MT is generated to be used in the simulation program.

3.3 Cache replacement policy

In this work, we consider I1 cache locking. A better cache replacement policy is adopted for I1. Using the MT, this policy always selects unlocked blocks with a minimum number of misses. In case of a tie in the number of misses, a block is selected randomly. However, a random cache replacement policy is used in D1 and CL2.

3.4 Multicore cache locking algorithm

In this subsection, we present the algorithm of the proposed multicore way locking technique. According to this algorithm, incoming tasks (i.e., requests) come to the controller of the multicore device. The controller assigns the incoming task to a free core (if any) and marks the core as busy. If no free core is found, incoming tasks are put into a waiting queue and served when cores become available (i.e., free) on a first come first serve basis. Depending on the assigned task to a core, the respective MT is populated, the pre-selected blocks are loaded, and a portion of the cache (I1 in this work) is locked. Then the core starts processing the task. In case of a cache miss, it selects the victim block (as the cache is already full) using the MT and modified cache replacement policy. The selected victim block is an unlocked block with the minimum number of misses. After the current task is finished, the core is made free (i.e., available) for performing another task.

4. Evaluation

We develop a simulation platform to evaluate our proposed multicore processor architecture. First, we briefly discuss the simulation details. Then, we evaluate our work by presenting some important simulation results.

4.1 Simulation details

We simulate the multicore architecture and cache locking using Moving Picture Experts Group's MPEG4 (part-2) decoding and Fast Fourier Transform (FFT) workloads. The code size for MPEG4 and FFT are 29,937 Bytes and 2,335 Bytes, respectively. We generate a tree graph for each application that shows the blocks that cause misses by single-core WCET analysis. By post-processing the tree graph, we create an MT for each application. We use VisualSim to develop the simulation platform, run the simulation program, and collect the results. In this subsection, we briefly discuss some assumptions, simulation tools, and important input and output parameters.

4.1.1 Assumptions

We make the following important assumptions,

- I1 way cache locking is implemented in this work.
- A modified cache replacement strategy is used for I1 using MT information.
- Write-back memory update policy is used.
- The (memory latency) delay introduced by the bus that connects CL2 and the main memory (Bus2 in Figure 1) is 15 times longer than the delay introduced by the bus that connects CL1 and CL2 (Bus1 in Figure 1).

4.1.2 Simulation tools

VisualSim and Heptane simulation tools are used in this work. VisualSim (a.k.a., VisualSim Architect) from Mirabilis Design is a graphical system-level simulation tool (VisualSim, 2022). We install VisualSim in Windows XP on a Dell PowerEdge 1600SC PC. Using VisualSim, we model the abstracted architecture. Using VisualSim simulation cockpit, we run the simulation program and obtain simulation results. The results are stored as text and/or graph files.

Heptane (Hades Embedded Processor Timing ANalyzEr) is a WCET analysis tool for embedded systems (Heptane, 2022). We configure Heptane in Red Hat Linux 9 in the same Dell PowerEdge 1600SC PC (where we have VisualSim in Windows XP). Once the configuration file is created, Heptane is run by typing “heptane-run.sh”' in a command shell provided that the PATH variable contains the directory where Heptane is installed. The results are placed in the directory as specified in the configuration file and are viewed through a Web browser by opening the file “HTML/index.html”.

4.1.3 Input and output parameters

Important input parameters are shown in Table 1. We simulate an 8-core system with I1, D1, and

CL2. We keep CL2 size fixed at 128 KB and vary the I1/D1 size, CL1/CL2 line size, and associativity level. Output parameters include average delay per task and total power consumption.

Table 1: Important Input Parameters

Parameter	Value
I1 (/D1) cache size (KB)	4, 8, 16, 32, or 64
CL1/CL2 line size (Byte)	16, 32, 64, 128, or 256
CL1/CL2 associativity level	1-, 2-, 4-, 8-, or 16-way
CL2 cache size (KB)	128 (fixed)
Number of cores	8 (fixed)

4.2 Results

In this work, we propose a multicore architecture with miss tables in CL1s to enhance the predictability and performance of power-aware visual computing systems. We model a system with 8 processing cores and run the simulation program using MPEG4 decoding and FFT algorithms. We obtain results by varying the I1 (/D1) cache size, I1 (D1/CL2) line size, and I1 (D1/CL2) associativity level with and without applying I1 cache locking. Based on our previous work (Asaduzzaman et al., 2006; Asaduzzaman et al., 2013), we apply 20% cache locking in this work for optimal predictability. In the following subsections, we discuss the impact of I1/D1 cache size, CL1/CL2 line size, and CL1/CL2 associativity level on average delay per task and total power consumption.

4.2.1 I1/D1 Cache size

We obtain the average delay per task for various I1/D1 cache sizes for no locking and 20% I1 locking using MPEG4 decoding and FFT workload. As shown in Figure 5, for any I1 cache size used, the mean delay per task for both MPEG4 decoding and FFT workloads decreases when I1 cache is locked partially. Results also show that the average delay per task decreases with the increase in I1 cache size.

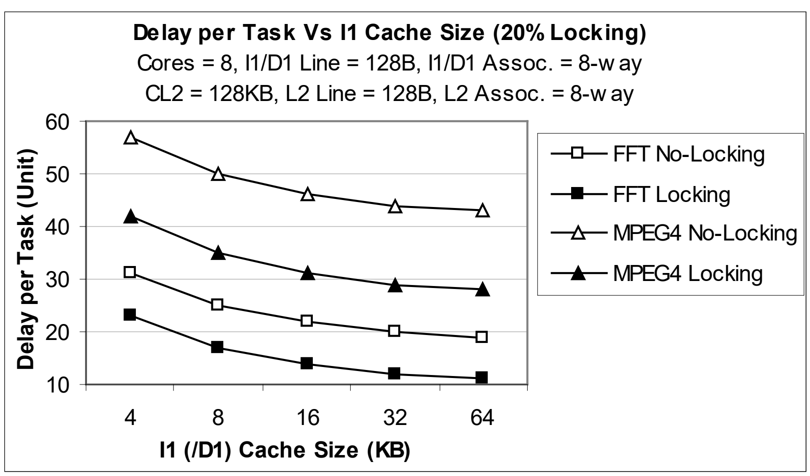


Figure 5: Mean delay per task versus I1 cache size

Experimental results show that regardless of the I1 cache size, total power consumption due to both MPEG4 decoding and FFT decreases when we compare no locking with I1 way locking [Figure 6]. Again, total power consumption decreases with the increase in I1 cache size. The decrease in total power consumption is significant for smaller I1 (4KB to 16KB).

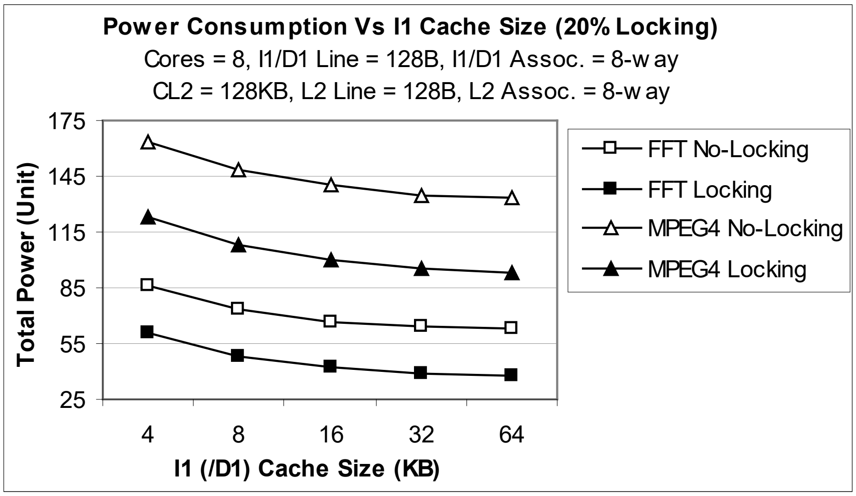


Figure 6: Total power consumption versus I1 cache size

4.2.2 L1/CL2 Line size

We use the same line size for I1, D1, and CL2. Figure 7 shows the average delay per task versus (I1) line size for no locking and 20% I1 way locking. We observe that the average delay per task goes down for MPEG4 and FFT workload, regardless of the line size, with increasing line size leveling off at 128B. For a line size greater than 128B, the average delay per task increases due to cache pollution.

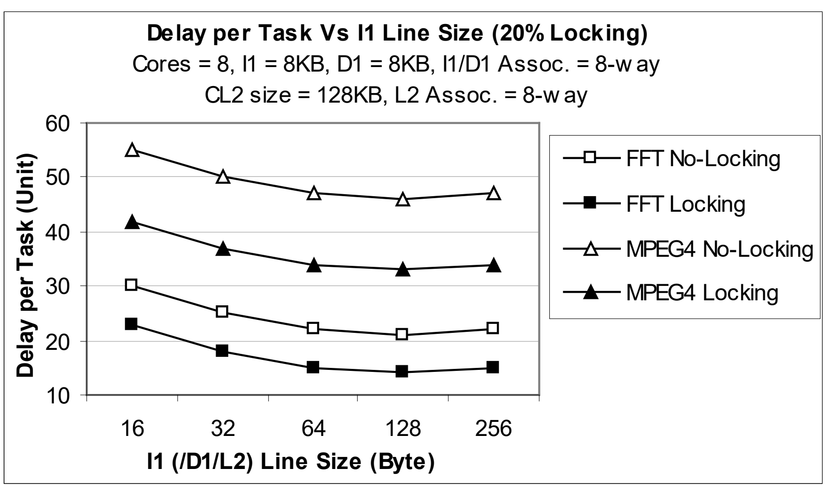


Figure 7: Mean delay per task versus I1 line size

Simulation results also show that total power consumption goes down for MPEG4 and FFT, regardless of the line size, with increasing line size leveling off at a line size of 128B (see Figure 8). It is also observed that for line size greater than 128B, the total power consumption increases.

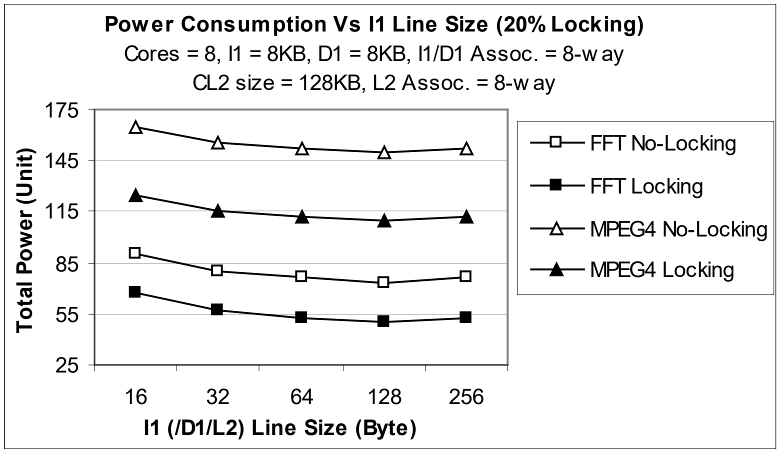


Figure 8: Total power consumption versus I1 line size

4.2.3 CL1/CL2 Associativity level

We use the same associativity level for I1, D1, and CL2. The impact of no locking and 20% I1 way locking on the mean delay and total power consumption by varying (I1) associativity level is shown in Figure 9. Experimental results show that for any I1 associativity level, the mean delay per task for both MPEG4 and FFT decreases when we move from no locking to 20% I1 way locking. Also, the decrease in the mean delay per task is significant for smaller levels of associativity (1-way to 2-way), after 4-way the mean delay per task remains almost the same.

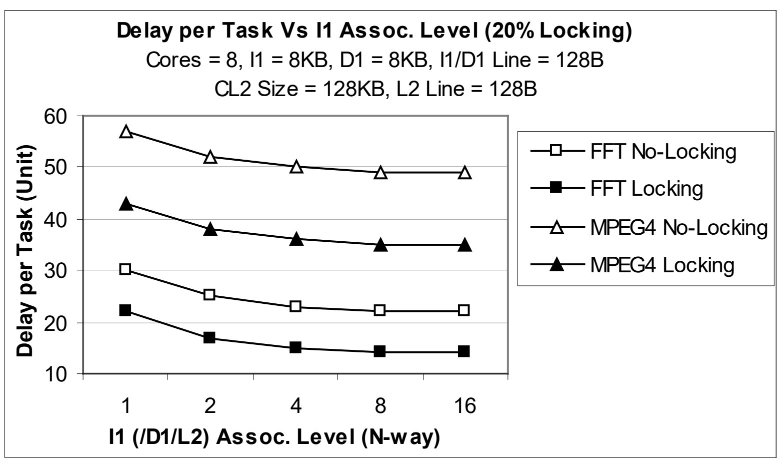


Figure 9: Mean delay per task versus I1 associativity level

We present the total power consumption due to different (I1) associativity levels for no locking and I1 way-locking in Figure 10. Simulation results show that regardless of the (I1) associativity level, the total power consumption due to both MPEG4 and FFT decreases when we compare no locking with 20% I1 way locking. Results also show that total power consumption decreases with the increase in (I1) associativity level. However, the decrease in total power consumption is significant between 1-way and 4-way, after that total power consumption remains almost the same.

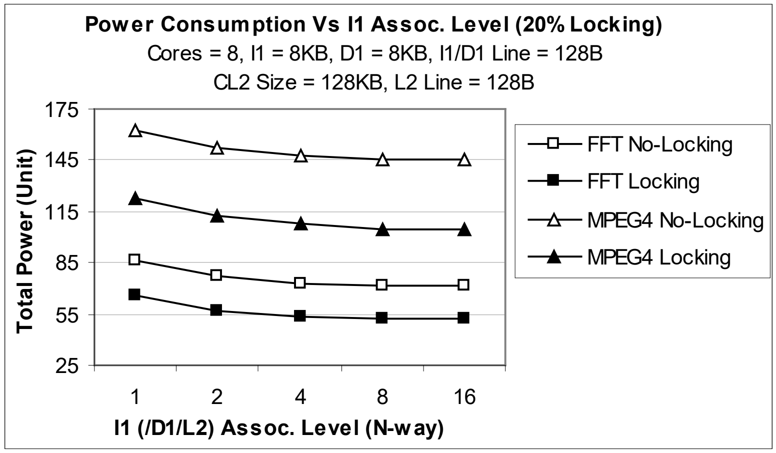


Figure 10: Total power consumption versus I1 associativity level

Finally, we summarize the changes in mean delay and total power consumption. The decrement in value is represented with leading (-) and the increment in value is represented with leading (+). As shown in Table 2, both the mean delay per task and total power consumption decreases when 20% I1 cache locking is applied in a proposed 8-core processor with MTs in CL1s. For I1 = D1 = 64KB, the mean delay per task is decreased by 42% and total power consumption is decreased by 40% for FFT (see Table 2, Figure 5, and Figure 6).

Table 2: Changes Due to 20% I1 Cache Locking

Parameters (Values)	Application Name	Changes (%)	
		Delay	Power
I1 Cache size (64 KB)	MPEG4	(-)35	(-)30
	FFT	(-)42	(-)40
I1 Line size (256 Byte)	MPEG4	(-)28	(-)26
	FFT	(-)32	(-)30
I1 Assoc. level (16-way)	MPEG4	(-)29	(-)27
	FFT	(-)36	(-)28

5. Conclusion

Traditional single-core processors are not capable of supporting future audio-visual computing. Multicore architecture provides a new platform for implementing highly complex audio-visual computing algorithms. The algorithms are becoming very complicated to solve the forthcoming problems efficiently. Complex multicore processors with multilevel caches are being deployed in both desktop and embedded computing to achieve a high performance-to-power ratio. Caches in multicore architecture make the execution time unpredictability worse. Therefore, running real-time applications on multicore visual systems becomes a serious challenge. In this work, we present a multicore processor with miss tables inside level-1 caches to increase predictability and performance-to-power ratio by cache locking. Miss table holds the block address information regarding the current applications being processed that causes cache misses. This approach reduces cache misses by locking/replacing the right cache blocks inside the cache which boost predictability and reduces average delay per task and total power consumption. We evaluate our proposed architecture by simulating an 8-core processor that has 2 levels of caches using MPEG4 decoding and FFT workload.

We find the proposed multicore processor with miss tables very promising. Experimental results show that a reduction of 35% for MPEG4 decoding and 42% for FFT in mean delay per task and a reduction of 30% for MPEG4 decoding and 40% for FFT in total power consumption are achieved by locking 20% of the total I1 cache size. Using the miss table information, better use of cache is possible to improve predictability and performance to power ratio. We plan to investigate the impact of miss table inside CL2 as an extension of this work in the near future.

6. References

Arnaud, A., Puaut, I. (2005). Dynamic Instruction Cache Locking in Hard Real-Time Systems. Retrieved on Oct. 30, 2022, from https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.332.4514

Asaduzzaman, A., Limbachiya, N., Mahgoub, I., Sibai, F. (2007, August 18-20). *Evaluation of I-Cache Locking Technique for Real-Time Embedded Systems*. In 2007 Innovations in Information Technologies (IIT), Dubai, United Arab Emirates.

Asaduzzaman, A., Mahgoub, I., (2006). Cache Modeling and Optimization for Portable Devices Running MPEG-4 Video Decoder. *Journal of Multimedia Tools and Applications (MTAP)*, 28(1), 239-256.

Asaduzzaman, A., Suryanarayana, V.R., Sibai, F.N. (2013). On level-1 cache locking for high-performance low-power real-time multicore systems. *Journal of Computers & Electrical Engineering*, 39(4), 1333-1345.

Blachford, N. (2005). Cell Architecture Explained Version 2. Retrieved on Oct. 30, 2022, from http://www.blachford.info/computer/Cell/Cell0_v2.html

Burt, J. (2009). IBM Says Cell Processor Technology Will Continue. Retrieved on Oct. 30, 2022, from https://www.eweek.com/networking/ibm-says-cell-processor-technology-will-continue/

Campoy, A.M., Tamura, E., Saez, S., Rodriguez, F., Busquets Mataix, J.V. (2005). *On Using Locking Caches in Embedded Real-Time Systems*. In: Yang, L.T., Zhou, X., Zhao, W., Wu, Z., Zhu, Y., Lin, M. (eds) Embedded Software and Systems. International Conference on Embedded Software and Systems (ICESS) 2005. Lecture Notes in Computer Science, 3820, 150-159. Springer, Berlin, Heidelberg.

Chiappetta, M. (2007). AMD Barcelona Architecture Launch: Native Quad-Core. Retrieved on Oct. 30, 2022, from https://hothardware.com/reviews/amd-barcelona-architecture-launch-native-quadcore

Heptane (2022). Online: inria. Heptane software - A tree-based Worst-Case Execution Time (WCET) analysis tool. Retrieved on Oct. 30, 2022, from https://team.inria.fr/pacap/software/heptane/

Inkley, B. and Tetrick, S. (2006). Intel Multi--core Architecture and Implementations. Retrieved on Oct. 30, 2022, from https://en.wikichip.org/w/images/c/c1/Multicore_MATS002_999_pct.pdf

Lutkevich, B. (2020). *Cache Memory*. TechTarget. Retrieved on Oct. 30, 2022, from https://www.techtarget.com/searchstorage/definition/cache-memory

Multi-core computing: Xeon and Athlon. (2022-July 30). In Wikipedia. http://en.wikipedia.org/wiki/Xeon , https://en.wikipedia.org/wiki/Athlon

Pentium 4. (2022). Online: hardware secrets. Inside Pentium 4 Architecture. Retrieved on Oct. 30, 2022, from https://hardwaresecrets.com/inside-pentium-4-architecture/

Puaut, I. (2006). Cache Analysis Vs Static Cache Locking for Schedulability Analysis in Multitasking RealTime Systems. Retrieved on Oct. 30, 2022, from https://www.cs.york.ac.uk/rts/wcet2002/papers/puaut.pdf

Sibai, F.N. (2008). On the Performance Benefits of Sharing and Privatizing Second and Third Level Cache Memories in Homogeneous Multi-Core Architectures. *Microprocessors and Microsystems*, 32(7), 405-412.

Suhendra, V. and Mitra, T. (2008). Exploring Locking & Partitioning for Predictable Shared Caches on Multi-Cores. In 2008 45th ACM/IEEE Design Automation Conference, Anaheim, CA, USA.

Tamura, E., Busquets-Mataix, J. V., Martín, J.J.S., Campoy, A.M. (2005). A Comparison of Three Genetic Algorithms for Locking-Cache Contents Selection in Real-Time Systems. In: Ribeiro, B., Albrecht, R.F., Dobnikar, A., Pearson, D.W., Steele, N.C. (eds) Adaptive and Natural Computing Algorithms. Springer, Vienna. https://doi.org/10.1007/3-211-27389-1_111

Tamura, E., Rodriguez, F., Busquets-Mataix, J.V., Campoy, A.M. (2004, June). High-Performance Memory Architectures with Dynamic Locking Cache for Real-Time Systems. In the Proceedings of the 16th Euromicro Conference on Real-Time Systems, 1-4, Italy

Tarui, T., Nakagawa, T., Ido, N., Asaie, M., Sugie, M. (1992). Evaluation of the lock mechanism in a snooping cache. In ACM Proceedings of the 6th international conference on Supercomputing. https://dl.acm.org/doi/pdf/10.1145/143369.143384

Thompson. (2012). Scratchpad memory: Shared L2 vs Cell local stores. Retrieved on Oct. 30, 2022, from https://scratchpaddmemory.blogspot.com/2012/05/shared-l2-vs-cell-local-stores.html

Vance, A. (2006-January, 22). Cell processor goes commando. The Register. Retrieved on Oct. 30, 2022, from http://www.theregister.co.uk/2006/01/22/cell_mecury_army/

Vera, X., Lisper, B., Xue, J. (2003). Data Cache Locking for Higher Program Predictability. *ACM SIGMETRICS Performance Evaluation Review*, 31(1), 272–282. https://doi.org/10.1145/885651.781062

Visual Computing (2017). Online: techopedia. Retrieved on Oct. 30, 2022, from https://www.techopedia.com/definition/16286/visual-computing

VisualSim (2022). Online: mirabilisdesign. VisualSim - A system-level simulator from Mirabilis Design, Inc. Retrieved on Oct. 30, 2022, from https://www.mirabilisdesign.com/visualsim/

7. Appendix

7.1 Appendix A: VisualSim Block Diagram

Figure A shows the VisualSim model of the proposed 8-core processor. In VisualSim, a system to be evaluated is described in three parts – Architecture, Behavior, and Workload. The architecture includes elements like processing core, cache, bus, and main memory. Behavior describes the actions performed on the system. Examples include network traffic shaping. The workload is the transactions that traverse the system such as network traffic.

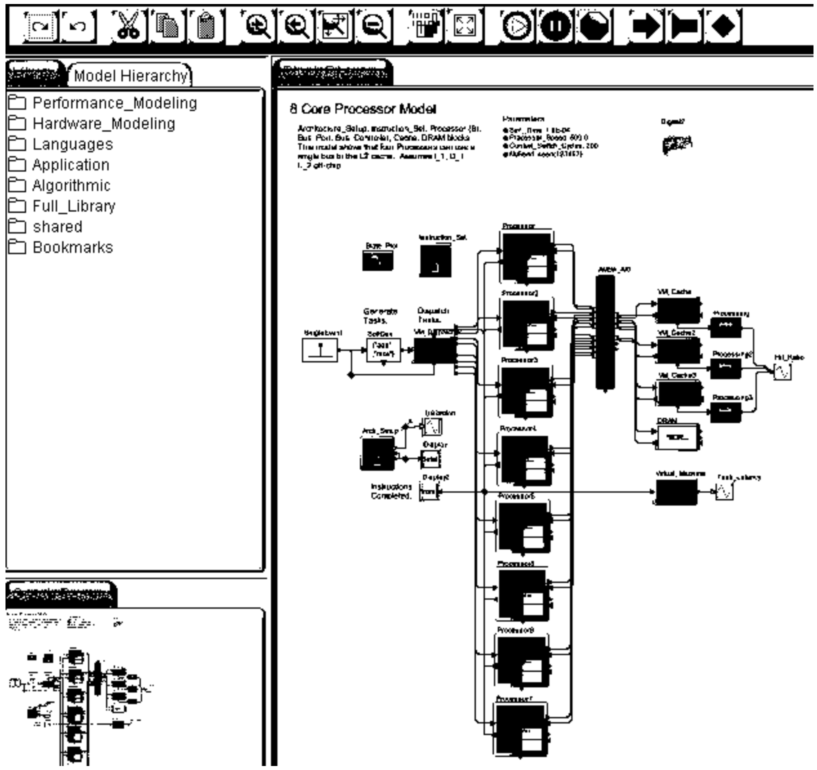


Figure A: 8-core processor model

7.2 Appendix B: VisualSim Simulation Cockpit

Figure B shows the VisualSim Simulation Cockpit. The Simulation Cockpit provides functionalities (left-top in Figure B) to run the model and to collect simulation results (right side in Figure B). Parameters can be changed before running the simulation without modifying the block diagram. The final results can be saved into a file and/or printed for further analysis.

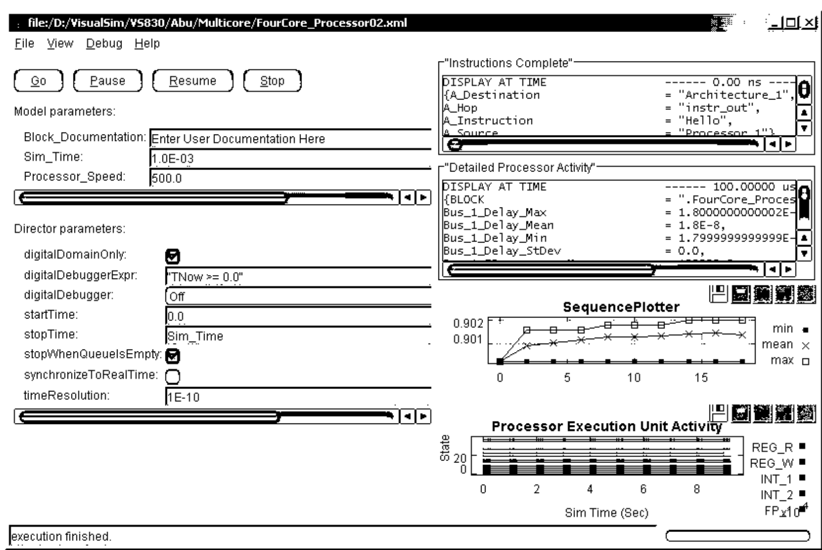


Figure B: VisualSim simulation cockpit for the model in Figure A

WINTER
2022

AIEMS
VOL 15
NO 2

AIEMS

AIEMS is committed to furthering the integration between business management, engineering, & industry. Our objective is to promote research, collaboration, & practice in these multidisciplinary areas. AIEMS seeks to encourage local, national, & international communication & networking via conferences & publications open to those in both academia & industry. We strive to advance professional interaction & lifelong learning via human & technological resources, and to influence and promote the recruitment and retention of young faculty and industralists.