Augmented Intelligence for Quality Control of Manual Assembly Processes using Industrial Wearable Systems

Short Paper

Introduction

Recent advances in sensor technology, a continuing decline of hardware prices and ubiquitous networking capabilities have led to a significant growth in Internet of Things (IoT) devices and applications. Fueled by innovations in machine learning and artificial intelligence, these new IoT devices become a leading driver of the ongoing digital transformation and enable a plethora of autonomous systems (Gubbi et al. 2013; Patel et al. 2017). Driven by the digital transformation, an increasing number of tasks can be automated substituting human work and forcing workers to adapt to this changing environment. The impact of increasing automation has often been discussed controversially (David and Dorn 2013; Loebbecke and Picot 2015; Rajnai and Kocsis 2017) and is attracting significant media attention. Still, many tasks cannot be fully automated. A case in point are complex assembly processes which easily surpass motion capabilities of current robot generations (David 2015; Gibbs 2016; Pfeiffer 2016). In these settings digital transformation is not about automation but rather about assisting and improving human performance by means of smart IoT devices. As pointed out by Pan (2016) and Pavlou (2018), human-computer symbiosis, also referred to as augmented intelligence, has the potential to leverage the complementary strengths of humans and computers.

However, there is no one-size-fits-all solution to develop and implement augmented intelligence systems. As smart IoT devices have to be newly developed or at least redesigned for many use-cases, the unique combination of hardware (sensors, motors, signals) and data processing during highly specialized processes will most of the time limit the direct applicability of existing training data or pre-trained machine learning models.

By means of a use case from the manufacturing sector we illustrate the bottom-up development process of an augmented intelligence system and highlight the important steps as well as the obstacles. Specifically, we design a wearable device for real-time quality control in an electronics assembly production step. Our example applications seeks to detect if connector systems (plugs) are properly connected during a manual assembly process. Driven by the "Poka-Yoke" principle, manufacturing companies strive to design fail-safe production processes (Dvorak 1998). It is for this reason that connector systems mechanically emit a distinctive acoustic signal ("click") to signify successful connections. However, such connections often have to be made under aggravated circumstances and outside a worker's line of sight (e.g., plugs have to be connected behind the glove compartment or in the drivetrain) while loud ambient noises overpower the click sound. Consequently, neither visual nor acoustic Poka-Yoke solutions are applicable. One way to overcome this obstacle is to augment the worker by means of a multi-use structure-born noise sensor which can detect object vibrations beyond superhuman levels. Such sensors can be embedded in a wearable device which is positioned at the workers' wrist (close enough for reliable detection, not impairing assembly motions). The device can then continuously record a broad band of frequencies transmitted via air (acoustic signals) or vibrations (structure-born noise). This hardware needs to be paired with an analytic backend that identifies valid click sounds from the sensor data stream. In turn the system can offer direct feedback with respect to the success of connections. Figure 1 illustrates a prototype of this IoT device.

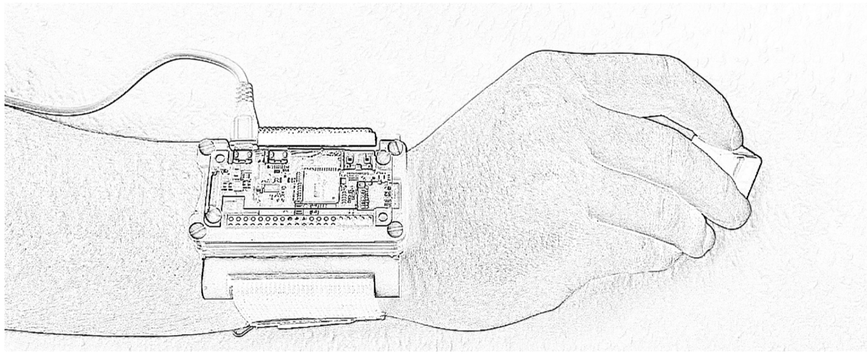

Figure 1. Prototype of the Wearable IoT Device

Related Work

Already in the the early 2000s Stanford (2002) showcased the deployment of wearable computing solutions in industrial environments. He characterizes multiple wearable computing projects. One of these early projects was Boeing's pioneer work on wearable-computing-based guided assembly (Mizell 2001). By recording test results this wearable smart device helped mechanics working on planes during assembly. More recent work focuses on wearable solutions for industrial maintenance (Zheng et al. 2015) which enables an efficient way for collaborating with a remote expert technology. Another approach builds a wearable devices securing safety system for the mining industry (Mardonova and Choi 2018). In particular they are using a sensor-equipped smart safety vest, eye-wear and smart helmet for real-time measurements of environmental conditions and the miner's health. Not putting special focus on wearables, Stein et al. (2018) highlight to what extent manual leakage detection for vacuum injection molds can be assisted by sensor data and spatial prediction models—another instance of augmenting worker capabilities by means of intelligent systems. Kong et al. (2018b) classify such systems as Industrial Wearable Systems (IWS). All IWS adopt a common design blueprint: data collection from equipment, human–machine cooperation, intervention and control of equipment.

In order to set up our system, we need to integrate our wearable computing sensor and a model for classifying the recorded stream. This model not only needs to identify the attributes of the specific noise but also distinguish it in different environments. The literature discusses different approaches for sound classification. Hidden Markov Models (HMM) have been the classic approach to classify acoustic signals (Su et al. 2011). These models allow to map individual parts of the sounds to states and compare them. More recently, machine learning approaches such as random forest tree classifiers (RF) (Saki and Kehtarnavaz 2014) and Support Verctor Machines (SVM) (Wang et al. 2006) are successfully applied for this task. With the rise of deep learning more recent studies focus on sound classification through Recurrent Neural Networks (RNN) (Vu and Wang 2016). RNNs are mainly used in the field of speech recognition which is closely related to noise detection (Graves et al. 2013). Finally, most recent work discusses the classification by utilizing deep convolutional neural networks (CNNs) (Li et al. 2018; Salamon and Bello 2017; Zhu et al. 2018). According to Mesaros et al. (2018) CNNs form the best basis for the classification of sound. The cited publication represents the "IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events"[1] (DCASE) 2018 challenge. Since 2013, the DCASE competition aims to continuously support the development of computational scene and event analysis methods by comparing different approaches using common publicly available benchmarks (Giannoulis et al. 2013). It is a well-known challenge in the sound classification research community presenting state-of-the-art architectures.

Conceptual Approach

We seek to complement a wearable sensor equipment with a data analytics backend to establish a real-time quality control system. To this end, we follow the Design Science Research paradigm which puts forward the

[1]http://dcase.community

development of useful artifacts as the central research goal (Baskerville et al. 2018; Von Alan et al. 2004). Such artifacts can either embody (i) new solutions for known problems, (ii) known solutions extended to new problems, (iii) new solutions for new problems, or (iv) known solutions for a known problem (Gregor and Hevner 2013). Along these lines our artefact instantiates as a new solution for a known problem as we combine existing components from different domains (information systems research, artificial intelligence) to a well known problem from quality control. Gregor and Hevner (2013) refer to such an artifact as improvement.

The structure-borne noise sensor combined with a Raspberry Pi module is worn on the worker's wrist without restricting the mobility. The device continuously streams sensor readings to a server using the Message Queuing Telemetry Transport protocol (MQTT), one of the standard IoT communication protocols (Al-Fuqaha et al. 2015). The predictive backend queries the data preparation module every second using the last five seconds of recording data. Based on the prepared data, the classification module provides real-time feedback to the worker. Our artifact is outlined in Figure 2.

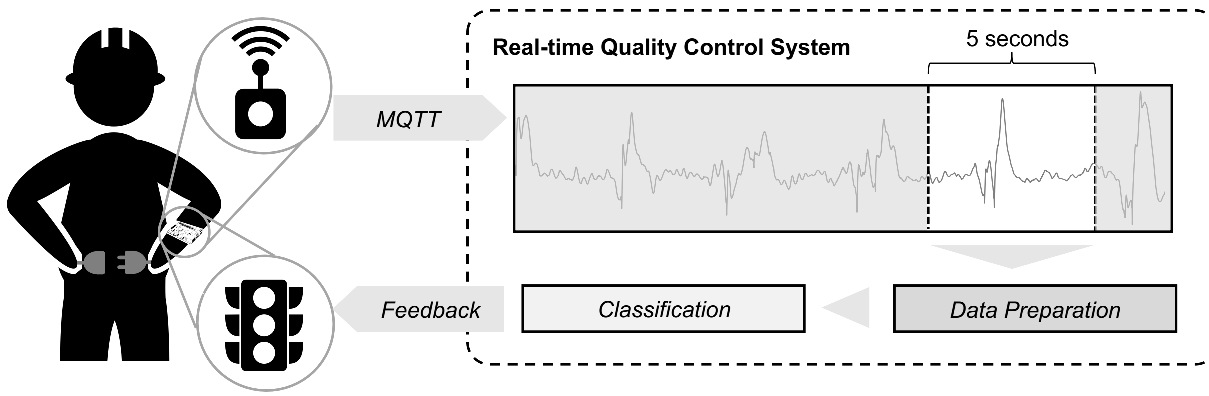

Figure 2. Artifact Outline

Experimental Design and Data Collection

Following Basili (1996), the quality and efficacy of a system has to be rigorously demonstrated by means of an appropriately selected evaluation method. While we aim to evaluate the artifact in a real production environment, we first have to show its viability. Hence, our initial case study relies on an experimental replication of the real-world assembly process.

To collect sufficient training data, we created a training program which repeatedly instructs the test person to perform one of the following actions in the next five seconds:

- Assemble the plug appropriately and thereby generate a positive example.
- Perform some kind of different movement and generate a negative example.

In order to ensure a similar distribution of the environmental sounds, the program randomly selects the action to be performed. Note that we opted for oversampling of negative examples as there is only one way to successfully connect the plugs but many ways to generate non-successful sounds (incomplete clicks, drops, walking, speaking, background noise).

Following this procedure, we collected a data set of 4,375 samples (1,525 positive and 2,850 negative). Each 5 second sample comprises an array of 160,000 sensor readings as well as a binary label (positive or negative). Figure 3 visualizes two examples of the raw data. In the right panel, a "click" is located between the two vertical dotted lines. Comparing the two samples it becomes obvious that there is a lot of noise in the data and that the correct assembly of the connector systems cannot readily be identified from raw data.

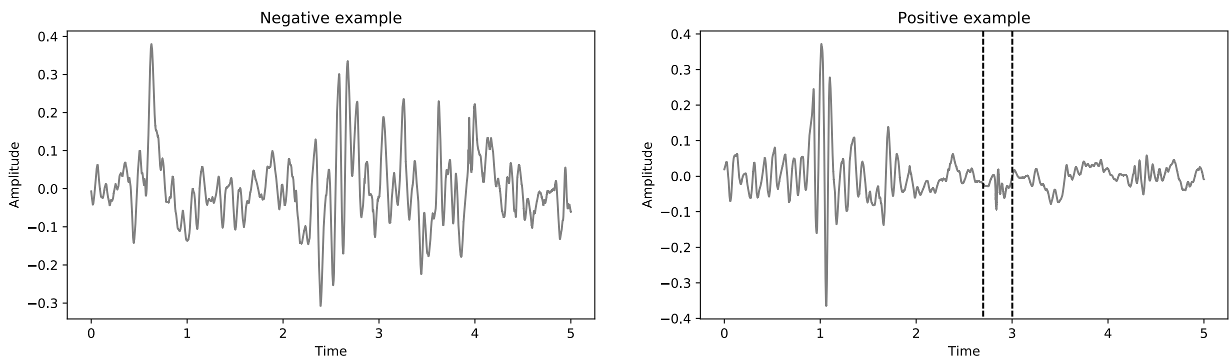

Figure 3. Raw Data

Data Preparation and Modelling

We apply a deep convolutional neural network (CNN) to classify whether or not a given sound sequence corresponds to a correct assembly of the plugs. In-line with Agarwal and Dhar (2014)'s call to action we primarily focus on problems and outcomes while limiting development efforts for new algorithms. Thereby, we follow (Griebel et al. 2019) and do not design a new network architectures from scratch but select one from state-of-the-art research papers solving similar problems.

Data Preparation

Even though CNNs render the task of manual feature engineering obsolete, the raw data still needs to be transformed in order to effectively train meaningful models.

On the one hand, network architecture for sound classification are designed to classify an acoustic signal based on its frequency spectrum. To obtain this, we decompose each recorded five second time window into its individual frequencies by means of the short-time Fourier transformation (Sejdić et al. 2009). This transformation splits a function of time (the sensor readings) into its frequencies (Bracewell and Bracewell 1986). Performing the Fourier Transformation on our one dimensional raw sensor data returns a two dimensional spectrogram. On the other hand, neural networks converge faster and therefore perform better if the input variables follow a standard normal distribution (LeCun et al. 2012). Hence, we perform a log transformation on the spectrogram and subsequently standardize the input variables. Figure 4 shows the data preparation pipeline on a negative as well as on a positive example.

Modeling and Training

As stated above, DCASE provides best practice models for sound classification. Therefore, we adopt the current DCASE-19 baseline model[2] (Kong et al. 2018a), which proved to be successful in the 2016 DCASE challenge (Valenti et al. 2017), to tackle the classification problem at hand. This CNN comprises two convolutional layers and one dense layer, followed by a sigmoid binary classification layer. For regularization we included batch normalization (Ioffe and Szegedy 2015) after each convolutional layer and dropout (Srivastava et al. 2014) after all layers.

In order to avoid overfitting we split our data into a training set (3500 samples ~80%) and a test set (875 samples ~20%). This is done in a stratified manner, maintaining the ratio of positive and negative samples from the original data. We additionally draw a random validation set (350 samples) from the training data to monitor the performance during model training and tuning. We preserve the test set for the final evaluation.

To increase generalizability as well as training stability, data augmentation is commonly applied to train deep neural networks. For image recognition tasks, this involves random transformations of each image

[2]https://github.com/qiuqiangkong/dcase2018_task1

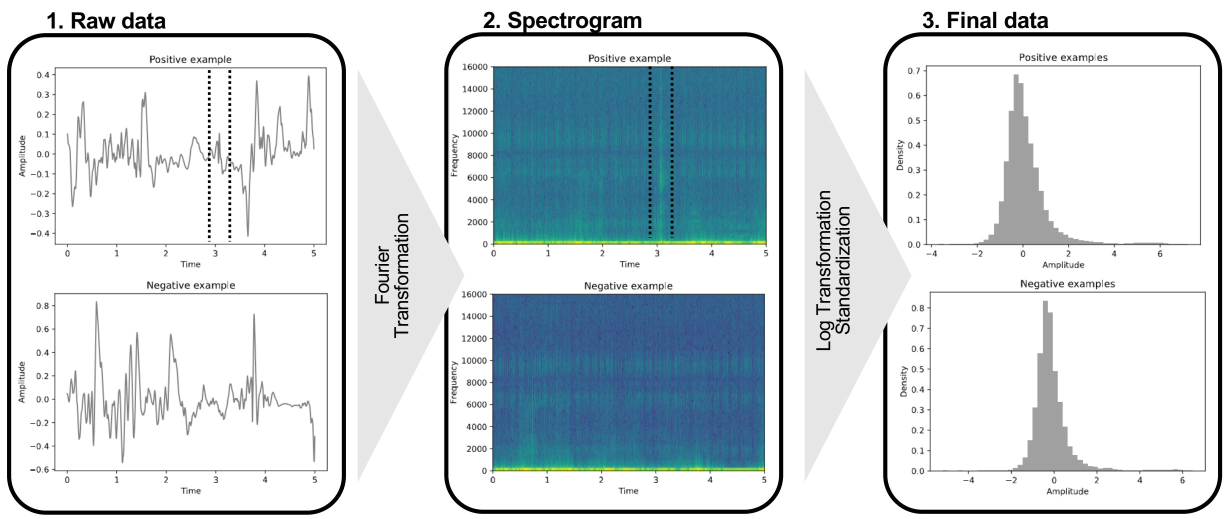

Figure 4. Data preparation pipeline

such as rotation, shearing, or flipping. In contrast to images, a spectrogram carries different information on each axis (i.e, frequencies, amplitude, and time). Hence, we can only apply transformations that do not change the sequence of the data. This renders the addition of Gaussian noise to each training sample as a valid remaining option for our case.

We implement the final model using the Tensorflow framework (Abadi et al. 2015). The training is performed on a Nvidia Tesla P100 GPU to minimize the binary cross entropy loss by means of the Adam optimizer (Kingma and Ba 2014).

Preliminary Results

We implement different state-of-the art audio classification approaches to assess the performance of our CNN. In contrast to deep neural networks, these models are based on hand-crafted features. Therefore, we extract 645 features from the spectrogram, namely the arithmetic mean, minimum, maximum, and median value for each frequency. We chose four different baseline models. These comprise two tree-based ensembles, a gradient tree boosting (XGB) (Chen and Guestrin 2016) and a random forest (RF) (Breiman 2001), as well as a support-vector machine (SVM) (Cortes and Vapnik 1995) and a Gaussian naive Bayes classifier (GNB) (Chan et al. 1982).

We chose the following evaluation metrics considering the class imbalance (more negative than positive samples) in our data set:

- *Matthews correlation coefficient* (MCC) is generally regarded as a good measure for imbalanced data (Powers 2011). It takes true positives (instances of correctly classified properly connected plugs), false positives (instances that contain falsely connected plug events but are erroneously classified as properly connected), true negatives (instances of falsely assembled plugs classified as falsely assembled plugs), and false negatives (instances of properly assembled plugs that are erroneously classified as falsely assembled) into account.
- *Precision* reports the fraction of correctly classified correctly assembled plugs among all instances that are classified as correctly assembled, i.e., true positives divided by the sum of true positives and false positives.
- *Recall* indicates the fraction of correctly assembled plugs that are correctly classified (true positives) among all correctly assembled plugs (true positives and false negatives).
- *F-Measure* considers both, precision and the recall, and is calculated as the harmonic mean of the two evaluation criteria.

Model	MCC	Precision	Recall	F-Measure
CNN	**98.74%**	99.67%	**98.69%**	**99.18%**
XGB	92.93%	96.32%	94.43%	95.36%
RF	90.98%	98.56%	89.51%	93.81%
SVM	76.58%	**100.00%**	68.52%	81.32%
GNB	25.22%	39.12%	99.67%	56.19%

Table 1. Classification Results on the Test set

As depicted in Table 1 the CNN achieves the best overall performance with an MCC of 98.74%, surpassing the second best model (XGB) by 5.81%. Notably, the SVM yields a precision of 100% (CNN 99.67%). It flawlessly classified all correctly assembled plug instances as correctly assembled. This can be particularly interesting for quality control systems that require high reliability. However, such systems should preferably yield a high recall as well. This holds true for the CNN, but not for the SVM.

Expected Contribution and Future Work

Our ultimate goal is to create an industrial wearable system to support quality control in connector systems assembly. By combining a hardware solution with a convolutional neural network we can perform real-time classification of assembled plugs based on structure-born noise signals. Our initial study yields promising results and establishes the feasibility of the suggested approach. However, our first findings are limited as the study was performed only on a single plug and on data collected by a limited number of test persons.

In light of our initial findings, we identify future research opportunities in various directions. Going forward, we plan to expand our test setting to more complex scenarios using different plugs and additional test persons. In particular, we want to quantify how much training data is needed to reliably adapt the neural network to changing conditions (i.e., different workers, different environments, different connector systems). This is of particular interest for companies planning to use such a system as test data collection is time-consuming and expensive. We posit that the additional data requirements can be limited by leveraging transfer learning principles which allow CNNs to efficiently adapt to related tasks.

Following Nunamaker Jr et al. (1990), evaluating the practical utility of the proposed artifact in the described experimental environment would already be a contribution to the research community. Going even further, the feasibility of the artifact should also be evaluated in a real-world assembly process (Sein et al. 2011).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. 2015. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.

Agarwal, R. and Dhar, V. 2014. "Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research* (25:3), pp. 443–448.

Basili, V. R. 1996. "The role of experimentation in software engineering: past, current, and future," in *Proceedings of IEEE 18th International Conference on Software Engineering,* IEEE, pp. 442–449.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., and Rossi, M. 2018. "Design science research contributions: finding a balance between artifact and theory," *Journal of the Association for Information Systems* (19:5), pp. 358–376.

Bracewell, R. N. and Bracewell, R. N. 1986. *The Fourier transform and its applications,* vol. 31999. McGraw-Hill New York.

Breiman, L. 2001. "Random forests," *Machine learning* (45:1), pp. 5–32.

Chan, T. F., Golub, G. H., and LeVeque, R. J. 1982. "Updating formulae and a pairwise algorithm for computing sample variances," in *COMPSTAT 1982 5th Symposium held at Toulouse 1982,* Springer, pp. 30–41.

Chen, T. and Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System," en. in. ACM Press, pp. 785–794.

Cortes, C. and Vapnik, V. 1995. "Support-vector networks," *Machine learning* (20:3), pp. 273–297.

David, H. 2015. "Why are there still so many jobs? The history and future of workplace automation," *Journal of economic perspectives* (29:3), pp. 3–30.

David, H. and Dorn, D. 2013. "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review* (103:5), pp. 1553–97.

Dvorak, P. 1998. "Poka-yoke designs make assemblies mistake-proof," *Machine design* (70:4), pp. 181–4.

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., and Ayyash, M. 2015. "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE communications surveys & tutorials* (17:4), pp. 2347–2376.

Giannoulis, D., Benetos, E., Stowell, D., Rossignol, M., Lagrange, M., and Plumbley, M. D. 2013. "Detection and classification of acoustic scenes and events: An IEEE AASP challenge," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* IEEE, pp. 1–4.

Gibbs, S. 2016. "Mercedes-Benz swaps robots for people on its assembly lines," *The Guardian* (26).

Graves, A., Mohamed, A.-r., and Hinton, G. 2013. "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on,* IEEE, pp. 6645–6649.

Gregor, S. and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS quarterly* (37:2), pp. 337–355.

Griebel, M., Dürr, A., and Stein, N. 2019. "Applied image recognition: guidelines for using deep learning models in practice," in *Proceedings of the 14th International Conference on Wirtschaftsinformatik,* pp. 393–406.

Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. 2013. "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future generation computer systems* (29:7), pp. 1645–1660.

Ioffe, S. and Szegedy, C. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *CoRR* (abs/1502.03167). arXiv: 1502.03167.

Kingma, D. P. and Ba, J. 2014. *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980.

Kong, Q., Iqbal, T., Xu, Y., Wang, W., and Plumbley, M. D. 2018a. "DCASE 2018 Challenge Surrey Cross-task convolutional neural network baseline," *Parameters* (4), pp. 4–691.

Kong, X. T., Luo, H., Huang, G. Q., and Yang, X. 2018b. "Industrial wearable system: the human-centric empowering technology in Industry 4.0," *Journal of Intelligent Manufacturing* (29:3), pp. 1–17.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. 2012. "Efficient backprop," in *Neural networks: Tricks of the trade,* Springer, pp. 9–48.

Li, M., Gao, Z., Zang, X., and Wang, X. 2018. "Environmental Noise Classification Using Convolution Neural Networks," in *Proceedings of the 2018 International Conference on Electronics and Electrical Engineering Technology,* ACM, pp. 182–185.

Loebbecke, C. and Picot, A. 2015. "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda," *The Journal of Strategic Information Systems* (24:3), pp. 149–157.

Mardonova, M. and Choi, Y. 2018. "Review of wearable device technology and its applications to the mining industry," *Energies* (11:3), p. 547.

Mesaros, A., Heittola, T., and Virtanen, T. 2018. "Acoustic scene classification: an overview of dcase 2017 challenge entries," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC),* IEEE, pp. 411–415.

Mizell, D. 2001. "Boeing's wire bundle assembly project," *Fundamentals of wearable computers and augmented reality* (5).

Nunamaker Jr, J. F., Chen, M., and Purdin, T. D. 1990. "Systems development in information systems research," *Journal of management information systems* (7:3), pp. 89–106.

Pan, Y. 2016. "Heading toward artificial intelligence 2.0," *Engineering* (2:4), pp. 409–413.

Patel, P., Ali, M. I., and Sheth, A. 2017. "On using the intelligent edge for IoT analytics," *IEEE Intelligent Systems* (32:5), pp. 64–69.

Pavlou, P. A. 2018. "Internet of Things – Will Humans be Replaced or Augmented?," *Marketing Intelligence Review* (10:2), pp. 42–47.

Pfeiffer, S. 2016. "Robots, Industry 4.0 and humans, or why assembly work is more than routine work," *Societies* (6:2), p. 16.

Powers, D. M. 2011. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies* (2:1), pp. 37–63.

Rajnai, Z. and Kocsis, I. 2017. "Labor market risks of industry 4.0, digitization, robots and AI," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY),* IEEE, pp. 343–346.

Saki, F. and Kehtarnavaz, N. 2014. "Background noise classification using random forest tree classifier for cochlear implant applications," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* IEEE, pp. 3591–3595.

Salamon, J. and Bello, J. P. 2017. "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters* (24:3), pp. 279–283.

Sein, M., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. "Action Design Research," *MIS Quarterly: Management Information Systems* (35:1), pp. 37–56.

Sejdić, E., Djurović, I., and Jiang, J. 2009. "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital signal processing* (19:1), pp. 153–183.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research* (15), pp. 1929–1958.

Stanford, V. 2002. "Wearable computing goes live in industry," *IEEE pervasive computing* (1:4), pp. 14–19.

Stein, N., Meller, J., and Flath, C. M. 2018. "Big data on the shop-floor: sensor-based decision-support for manual processes," *Journal of Business Economics* (88:5), pp. 593–616.

Su, F., Yang, L., Lu, T., and Wang, G. 2011. "Environmental sound classification for scene recognition using local discriminant bases and HMM," in *Proceedings of the 19th ACM international conference on Multimedia,* ACM, pp. 1389–1392.

Valenti, M., Squartini, S., Diment, A., Parascandolo, G., and Virtanen, T. 2017. "A convolutional neural network approach for acoustic scene classification," in *2017 International Joint Conference on Neural Networks (IJCNN),* IEEE, pp. 1547–1554.

Von Alan, R. H., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS quarterly* (28:1), pp. 75–105.

Vu, T. H. and Wang, J.-C. 2016. "Acoustic scene and event recognition using recurrent neural networks," *Tech. Rep.* (2016).

Wang, J.-C., Wang, J.-F., He, K. W., and Hsu, C.-S. 2006. "Environmental Sound Classification using Hybrid SVM/KNN Classifier and MPEG-7 Audio Low-Level Descriptor," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings,* 2006, pp. 1731–1735.

Zheng, X. S., Matos da Silva, P., Foucault, C., Dasari, S., Yuan, M., and Goose, S. 2015. "Wearable solution for industrial maintenance," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems,* ACM, pp. 311–314.

Zhu, B., Xu, K., Wang, D., Zhang, L., Li, B., and Peng, Y. 2018. "Environmental Sound Classification Based on Multi-temporal Resolution Convolutional Neural Network Combining with Multi-level Features," in *Pacific Rim Conference on Multimedia,* Springer, pp. 528–537.