IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 9, Issue 4, Dec 2021

The Use of Supervised Machine Learning for Classification Purposes with a Method for Sensor Reduction

RANI DUBEYA

Abstract—

Rapid adoption of sensor-based feedback and control systems in smart gadgets. Markets that place a premium on affordability are among the most likely to embrace these devices. Conventional machine learning-based control systems often incorporate data from several sensors in order to achieve performance objectives. Another method is presented that uses the time series data collected by a single sensor. Domain experts' knowledge of the system's physical occurrences is used to segment the time series output into discrete time chunks. The machine learning system's characteristics are derived from statistical observations over many time periods. When more characteristics are found that decouple vital physical measurements, the system's performance is improved. This state-of-the-art approach requires fewer observations than conventional methods, yet produces equivalent precision. Because of the reduced number of sensors and the considerably streamlined and more robust algorithm development and testing stage, the resulting development effort is far more cost-effective than that of traditional sensor categorization systems. The authors present their conclusions by analyzing a case study of a media-type classification system used in a commercially available printing system.

I. INTRODUCTION

Sensors are rapidly decreasing in cost while performance and accuracy increase. Consequently, many electromechanical devices have incorporated sensor-enabled control schemes. Recently, machine Learning algorithms have begun to leverage this trend to enable new functionality. Sensed information may Be used to generate input features for algorithms that enable proactive diagnostics, system-awareness, and other more complex tasks such as classification. Concerns arise when the number of sensors and the capability of individual nodes are constrained due to cost or other associated factors like computation time and memory footprint. Previous efforts to address this concern have focused on a reduction of computational requirements during both the training and classification phases of embedded supervised machine learning algorithm development [1]. Methods attempting to minimize the number of features required for classification also exist; these may be used to reduce the number of sensors necessary for a given task

.

ASSISTANT PROFESSOR, Mtech,
Department of CSE
Gandhi Institute for Technology, Bhubaneswar.

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

This work presents a novel method to reduce the number of sensors required for a supervised machine learning classification system. Expert knowledge of expected sensor output variation as a function of intrinsic properties, extrinsic properties, and uncontrollable external factors is used to establish a unique feature set that sufficiently decouples otherwise inseparable classes. The system design and control system were concurrently tuned to elicit distinct dynamic responses within predefined temporal regions of a continuous data stream. The analog data was discredited into several distinct zones of interest corresponding to the sensors response to different dynamical processes. A unique difference method allowed the learning algorithm to extract additional useful information

From the confounded data set. This methodology is validated by a case study of a print media classifier system developed for a commercial laser printer, which was manufactured and deployed at a large volume. The resultant classification success exceeded that of embodiments using multiple sensors with only a single sensor. Finally, the implications of this design methodology and advantages over a traditional data-driven classification system are discussed.

II. BACKGROUND

The goal of simplification of multi-sensor systems by harvesting more independent features from a reduced sensor set relies on modification of the measured object usually based on time or geometry. There are numerous studied methods for dimensionality reduction and representation of time series data. General dimension-reduction and re-representation methods include model-based techniques such as those using hidden Markov models [2], [3]. A second class of methods have attempted to reformulate the data with interpolative or regression methods such as piecewise linear (PLA) [4] or piecewise polynomial (PPA) [5] approximations. Another group of methods uses a symbolic representation optimized with certain constraints such as symbolic aggregated approximation (SAX). Still other methods use transforms such as discrete Fourier [6] or discrete cosine transforms or wavelet systems [7], [8]. Although these methods are largely designed for use on general, potentially multi-dimensional time series, they are frequently tested, presented, and verified on application specific data from medical data [8] to faults in mechanical gear systems [9]. Once the transformation has been performed, classification training and evaluation can occur. Possible algorithms include 1-nearest neighbors (1NN) or k-nearest neighbors (kNN) [10], which demonstrated considerable success when implemented with representations like SAX in combination with dynamic time warping [11]. More sophisticated methods such as neural networks, multi-layer perceptions [12], Bayesian networks [13], support vector machines [14] and decision trees [15] have also been used with success and represent alternative design options. Some methods use information from a transformation, such as warping distance, as an additional feature and integrate this into the classification method [16]. In each case, the features used to train these systems are selected to be as orthogonal as possible and the quality of the resulting algorithm is, amongst other things, a function of that orthogonality. Often, the system cannot be easily simplified, and hardware with embedded supervised machine learning systems is designed using a complex network of various sensors. In theory, this

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

extra data enables the designer to build and test a robust algorithm since a network of sensors can be selected to maximize feature orthogonality. This can lead to a temptation to deploy more sensors and computational resources than is strictly necessary. In industries where customers are highly sensitive to product cost, such as office printing, the strategy is often to deploy a single sensor to partly meet design needs. These attempts have included using a set of electrodes to take electrical measurements of media [17], [18], a camera to measure surface roughness [19] or an ultrasonic sensor to determine media density [20].

III. METHODOLOGY

In concurrently developed physical systems, the designer has access to significantly more information about the situation than is often available with analyzing time series data in a general case. Time series data output by a single sensor may contain information about multiple physical quantities due to system dynamic behavior. Therefore, multiple physical quantities do not always need to be measured by the same number of physical sensors. The designer has an opportunity to tune the hardware to produce a time series output from a single sensor and then discredited the output with domain expert knowledge to produce multiple features while preserving orthogonality.

This results in a system with fewer sensor nodes and a lower associated cost. Consider the case of a least-squares support vector machine (LS-SVM) [21], [22] deployed in an embedded classification, solving a multiclass problem (e.g. determine if a presented set of features belongs to which one of several distinct sets).

The goal is to take as input a vector x 2 Rnf , where nf is the number of features used for classification, and produce an output y(x) which represents the classifier output. Given xk 2 Rnf; k = 1; 2;:::; N are the feature vectors corresponding to N training examples and yk are the corresponding true classes (in this case yk = +1 if the measurement belongs to a set and yl = 1 if it does not), the classification algorithm is trained by solving the following optimization problem to determine a best separating hyper surface defined through a nonlinear mapping. Some interpretations of the LSSVM and other SVMs make assumptions about the variables being independent and identically distributed random variables. While we cannot make this claim for this dataset due to temporal correlation, SVM-type algorithms can still work well in practice as long as the combination of features can

Provide sufficient separation. In the implementation section, we discuss the distribution of the selected input features, and it can be observationally inferred that an SVM might work well given a geometric rather than probabilistic interpretation of SVM methods.

$$\text{minimize } J_P(w,e) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2 \qquad (1)$$

$$\text{subject to } y_k[w^T \varphi(x_k)_b] = 1 - e_k, k = 1,\ldots,N \qquad (2)$$

Where the classifier takes the form: y(x) = sign [wet'(x) + b], and '(xk) is a mapping to a (often) higher dimensional

Space. In practice the classifier is usually solved for in the dual space, the space of Lagrange multipliers of the constraints, _k (for k = 1; 2; N). b is a scalar bias offset term. Is a regularization parameter that can be used to control over fitting vs. under-fitting behavior, but was set as 1? W 2 Ruff is a vector of weights that, along with the mapping '(xk) helps to

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

define the decision hyper surface. The dual space classifier takes the form:

$$y(x) = \text{sgn}[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b] \quad (3)$$

$K(x, x_k) = \varphi^T(x)\varphi(x_k)$ Is a Kernel function (a nonlinear mapping that allows additional flexibility in the classification function). Both the dual space classifier and the solution of the classifier optimization problem can be addressed by considering the Karsh-Kuhn-Tucker (KKT) conditions for optimality:

$$\mathbf{w} = \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k),$$
$$\sum_{k=1}^{N} \alpha_k y_k = 0,$$
$$\alpha_k = \gamma e_k, \forall k = 1, 2, ..., N,$$
$$y_k[\mathbf{w}^T \varphi(x_k) + b] - 1 + e_k = 0, \forall k = 1, 2, ..., N.$$

This allows assembly of the following matrix equation to solve the KKT system:

$$\begin{pmatrix} 0 & \mathbf{y}^T \\ \mathbf{y} & \mathbf{\Omega} + \frac{\mathbf{I}}{\gamma} \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 1_v \end{pmatrix} \quad (4)$$

where $\Omega_{kl} = y_l y_l \varphi(x_k)^T \varphi(x_k) = y_k y_l K(x_k, x_l)$, with $k, l =$ 1; N. At this point the (no sparse) matrix equation can be solved for _ and b using standard methods (LU factorization, etc.). The Kernel function can take a number of different Forms, of which

$K(x, x_k) = x_k^T x$ (linear), $K(x, x_k) =$ $(x_k^T x + \tau)^d$ (polynomial), and $K(x, x_k) = \exp\left(-\frac{\|x - x_k\|_2^2}{\sigma^2}\right)$

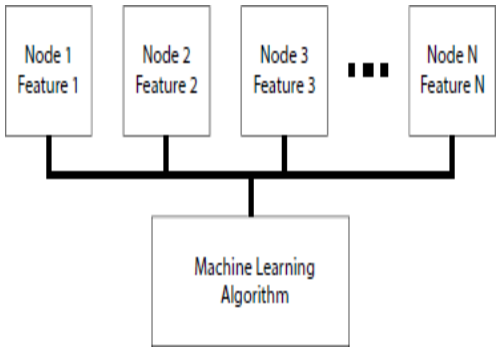


Fig. 1. Traditional method for enabling feature-based decision making capability on an existing device. The final classification algorithm is a function of N features, represented by N nodes. In this work, only polynomial classifiers are considered. This is due to the application requirements of processing power and program memory space, constrained to use the algorithm of [1]. Typically, y(x) = +1 would yield a prediction that x belongs in one set, and y(x) = 1 would correspond to the other complementary set. However, in some cases, including media classification in a printer, there are areas of the feature space that for some comparisons make no difference (there is cases in which mistakes in classification cause less of a problem for downstream processes). Specifically, one can have

Some errors in classification that is acceptable to downstream processes, and some that should be weighted more heavily. This idea was discussed and formulated into the training of a multiclass SVM problem and described in detail in [23]. The solution method for the system is the same. In this work, the result associated with each classification is accordingly either an incorrect classification, an incorrect (but acceptable) error, or a correct classification. An acceptable error is simply one that is tolerable to the downstream processes. In order to

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

create a multiclass classification system, the different classes are separated into complementary groups and evaluated in a one vs. all sense [22] (other options exist, but one vs. all is the encoding used in this work); if there are three

Classes, then there are three classifiers, each of which evaluates whether the data belongs in one set, or alternatively, all of the other sets. As mentioned before, selection of the features that comprise the feature vector are critical to classifier performance. The focus of this work is the design of the features and corresponding sensors and mechanical elements needed in order to achieve good performance while minimizing training data and overall cost.

A traditional approach, shown in Figure 1 places the burden of the system on the sensor nodes themselves. In this example, a feature contributing to the classifier has a one-to-one relationship to the number of required sensor nodes. The proposed approach illustrated by Figure 2 puts the burden of the system on the domain expert knowledge and the temporal output of a single node. The domain expertise is used to partition the measurement time series m (t) (in implementation, this is most likely a sampled time series) into discrete intervals, such that

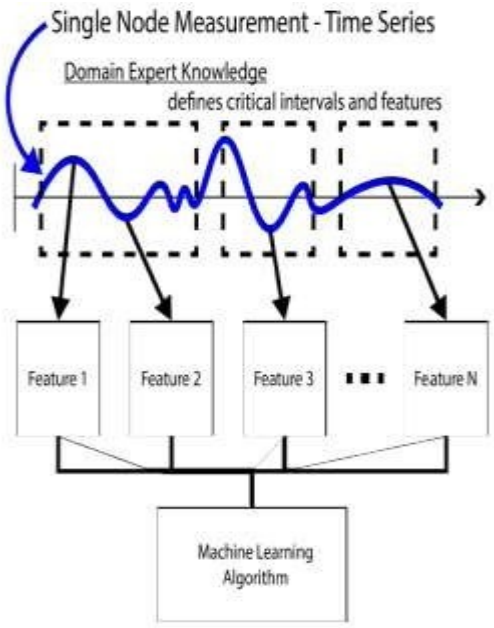


Fig. 2. The proposed approach uses the system knowledge of co-designed hardware to pull multiple features out of a single time series of data.

$$m(t) = \begin{matrix} [x(t_1,t_2): & [\Psi_{t_1,t_2}], \\ x(t_2,t_3): & [\Psi_{t_2,t_3}], \\ \vdots & \\ x(t_{N-1},t_N): & [\Psi_{t_{N-1},t_N}]] \end{matrix}$$

The classifier is trained on data that is of the form (YK, xk). Ideally, ski = _k, where _k is the set of intrinsic physical

Properties in the system (_k = [_1; _2; _Nape] To 2 Ropy).

Nape represents an ideal set of orthogonal intrinsic properties. _ _k. simply put, ideally, the sets to be classified are well separated by a measurement of some direct, relevant intrinsic physical property and have good orthogonality. In the practical case, this is not so. Every measurement is a function of both the intrinsic property being measured and the properties of the physical system involved in that measurement.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

These properties include the structure of the system and its operation, which are controllable by the system designer, and known environmental factors which may not be controllable by the designer. Considering the form of the constructed intervals and corresponding statistical measures, the training data examples ski are such that

$$x_k = [f_1(\phi_k, Y_1, Z_k), \\ f_2(\phi_k, Y_2, Z_k), \\ \vdots \\ f_N(\phi_k, Y_N, Z_k)]$$

Here, (f1; f2; fan) are nonlinear functions of the arguments: _k, the intrinsic physical properties; Ski 2 Rape

Which are known, quantifiable extrinsic system properties that influence the measurement (Npe is the number of extrinsic properties affecting measurements); and (Y1; Y2; YN), which are uncontrollable external factors that are a function of the hardware design.

In the case of systems where measurements taken in different intervals are coupled, taking the difference between

Two functions can help to train the classifier with independent information about system interactions and decouple external factors that influence the measurement. This can be justified with a brief expansion analysis. Given two functions if and fjord, the Taylor series expansions can be taken about a nominal operating point as

$$f_i(\phi_k, Y_i, Z_k) = \frac{\partial f_i}{\partial \phi_k}\Delta\phi_k + \frac{\partial f_i}{\partial Y_i}\Delta Y_i + \frac{\partial f_i}{\partial z_k}\Delta Z_k + C_i \quad (5)$$

$$f_j(\phi_k, Y_j, Z_k) = \frac{\partial f_j}{\partial \phi_k}\Delta\phi_k + \frac{\partial f_j}{\partial Y_j}\Delta Y_j + \frac{\partial f_j}{\partial z_k}\Delta Z_k + C_j \quad (6)$$

Taking the difference yields

$$f_i(\phi_k, Y_i, Z_k) - f_j(\phi_k, Y_j, Z_k) = \\ \left(\frac{\partial f_i}{\partial \phi_k}\Delta\phi_k + \frac{\partial f_i}{\partial Y_i}\Delta Y_i + \frac{\partial f_i}{\partial Z_k}\Delta Z_k + C_i\right) - \\ \left(\frac{\partial f_j}{\partial \phi_k}\Delta\phi_k + \frac{\partial f_j}{\partial Y_j}\Delta Y_j + \frac{\partial f_j}{\partial Z_k}\Delta Z_k - C_j\right) = \\ \underbrace{\Delta\phi_k\left(\frac{\partial f_i}{\partial \phi_k} - \frac{\partial f_j}{\partial \phi_k}\right)}_{\Delta\phi_k = 0 \text{ for same } k} + \frac{\partial f_i}{\partial Y_i}\Delta Y_i - \frac{\partial f_j}{\partial Y_j}\Delta Y_j + \\ \underbrace{\Delta Z_k\left(\frac{\partial f_i}{\partial Z_k} - \frac{\partial f_j}{\partial Z_k}\right)}_{\Delta Z_k = 0 \text{ for same } k} + \underbrace{C_i - C_j}_{\text{constant}}$$

For the same training example,_k = 0. The same is true for _Ski. Therefore, the only remaining terms are those that

Include _Yi and _I, the associated partial derivatives, and the difference of the offset constants. This new feature, fifj , is solely a function of _Yi and _Yj , which are functions of certain fixed extrinsic system properties. This information can be learned by the classifier and improve classification performance.

IV. CASE STUDY AND IMPLEMENTATION

This case study applies the proposed approach to a commercial color laser (electro photographic) printer intended for

Shared office use in a managed print services environment. Most laser printer users do not check or adjust the media

Type settings. Additionally, only a fraction of users that do adjust the media settings do so correctly. Incorrect settings on these devices may cause problems for both the customer and the manufacturer. To address this issue, an inexpensive sensor system and embedded machine learning algorithm were implemented to classify media without user input.

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

The printer control system adjusted device parameters based on this media classification.

A single inexpensive optical sensor consisting of a paired LED and phototransistor was mounted within the printer

Media path. The sensor output a continuous data stream corresponding to the amount of light transmitted by in-process media. A simple model of the sensor was developed and, based upon this, system hardware and controls were tuned to generate an information-rich data stream by leveraging the dynamic response of media to control system inputs. The printer generated features from this data stream for each sheet of media. A broad population of standard office media with varied intrinsic properties, 'k, existing along a continuum was sorted into 1 of 5 distinct classes: light, normal, heavy, card stock, and transparency. This dataset was used to generate an embedded machine learning algorithm that used these features to determine media class in near real time. Printer process parameters and system controls were adjusted based upon this prediction. The final embodiment significantly reduced overall cost, complexity, and system footprint when compared to traditional implementations and is described in greater detail in [24]. A cross section of the printer media path is shown in Figure 3. The highlighted region contains a section view of the sensor and the surrounding printer hardware including upstream feed rollers, media guides, and downstream feed rollers. The electrical design schematic for the optical sensor system is shown in Figure 4. Nominal circuit values were tuned to adjust the sensor gain, response, and sensitivity. The resulting full scale range of the data set was maximized for the population of expected media and maximum separation between media classes was achieved. Calibration was performed to compensate for system gain and offset errors. The sensor outputs a continuous data stream corresponding to the amount of infrared light transmitted by the in-process media. This output is a highly coupled function of many confounded factors including intrinsic media properties (e.g., media basis weight, media roughness, media thickness, etc.) 'K, extrinsic system properties (e.g., LED intensity, media speed, media input source, phototransistor sensitivity, feed roller velocities, media shape and offset, LED and phototransistor directionality, etc.) Zk and uncontrollable external factors (e.g., relative humidity, temperature, etc.) Yi. Figure 5 depicts how variability caused by these confounded factors impacts the measurement for a single media. 20 measurements for a normal weight office media are shown. The signal varies substantially from sheet to sheet and within a given sheet. Sensor output for a given media may vary as much as 20% of the sensors full scale range at a given process point. This is primarily a function of intrinsic media properties, 'k. Within a given sheet, the sensor output may vary as much as 60% of the sensors full scale range. This is primarily a function of extrinsic system properties, Zk. Uncontrollable external factors, Yi, alter both the intrinsic media properties,

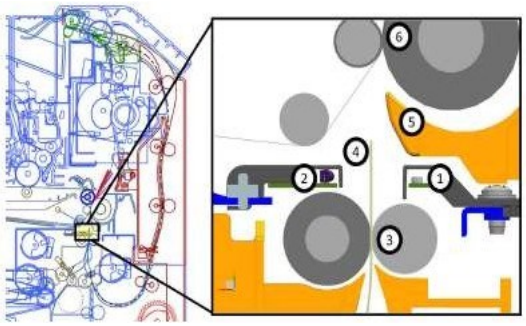


Fig. 3. A cross-section of the printer media path is depicted. The highlighted region contains an optical

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

sensor consisting of an LED (1) and phototransistor (2) that measures the amount of infrared light transmitted by a sheet of media (4) as it is processed by the printer. Media fed by upstream feed rollers (3) passes through the sensor, beyond a media guide (5), and into a set of downstream feed rollers (6). Hardware (physical design of the media path)

And firmware (system timings and relative velocities of the feed rollers) were tuned during development to enhance data orthogonality by controlling the position and shape of the media relative to the sensor in the spatial/temporal

Domain.

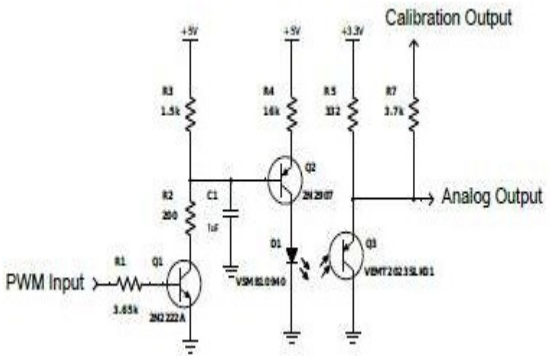


Fig. 4. The electrical design schematic for the optical sensor system is shown. The connection labeled “Analog Output” is the voltage signal measured by the analog-to-digital converter and used in the classification system. Nominal circuit values were selected to optimize the sensor gain, response, and sensitivity for a broad range of media types. and extrinsic system properties, Zk.

Figure 6 depicts how this variability manifests as boundary confusion. A broad set of standard office media possessing a range of intrinsic properties, 'k, existing along a continuum were used to train and test the algorithm and are listed in Table II for reference. Corner cases (distinguishing light from card stock, for example) are easily distinguished. However, media properties exist along a continuum and variability from sheet to sheet and within a given sheet made the classification problem particularly challenging. There was a large amount of boundary confusion. This is especially true for the heavy class of media which significantly overlaps with both the normal and

Card stock classes.

For a classifier to be successful, it must decouple the relevant intrinsic media properties, 'k, from the other confounding variables and generate a substantially orthogonal feature set. Media to media variability must be decoupled from the variability seen from sheet to sheet or within a given sheet. For the case of media classification, this was achieved

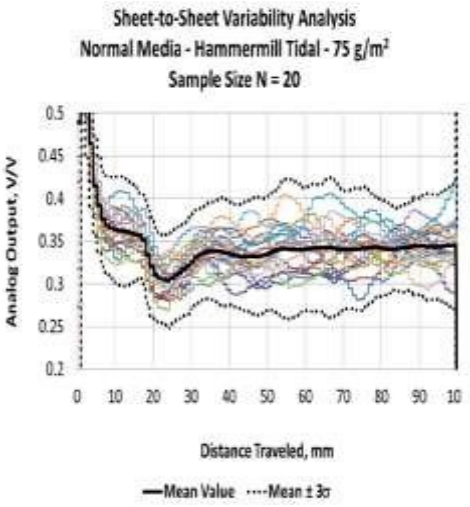


Fig. 5. Normalized analog sensor output for 20 separate sheets of a standard office paper are plotted. Data was collected for 100 millimeters of media travel. The population means and 99.7 percent confidence bands for this given

Media are plotted for reference.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

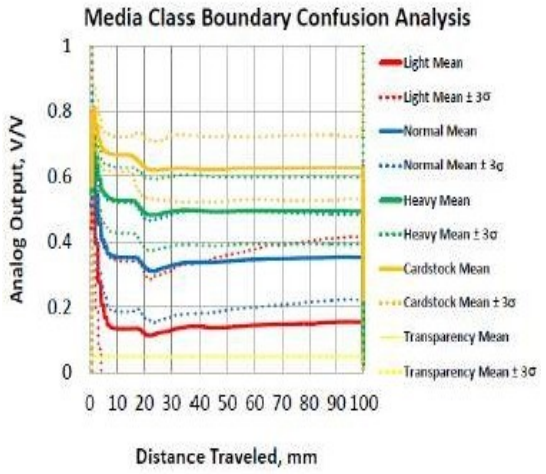


Fig. 6. Normalized analog sensor output for the mean and 99.7 percent confidence band of each class are plotted. The population for each class consists of 360 training samples from each media listed in Table II. A standard classification problem utilizing a traditional feature set would be intractable due to the continuous, overlapping nature of the data. By tuning system hardware and control parameters to leverage the sensitivity of the measurement to uncontrollable external factors, Yi, and extrinsic system properties, Zk. Since the sensor output was a nonlinear function of 'k, Zk, and Yi, it was possible to use the dynamic response of the system to help decouple these convoluted variables using the difference method described previously. Concurrently developed printer control algorithms and sensor hardware were tuned during the development phase to generate a continuous data stream that could be deconstructed into several distinct zones of interest corresponding to the sensors response to different dynamical processes. The resultant time series data was divided into 5 distinct zones of interest that corresponded to changes in the printer process that were designed to elicit a varied response from the sensor. In order to make the design more insensitive to printer-to printer variation, four ideas were considered when designing the zone positions. First, a flag sensor (integrated into the paper feed control system) allowed accurate registration of the leading edge of the sheet, and the traverse distance was known from the paper feed drive encoders. Second, the zones are larger than strictly necessary for a single printer in order to accommodate variation around the population of printers (determined empirically from a number of different printers).

It is important to be aware that performance can decrease if the buffer regions are too large as the data quality will decrease from the statistical measure being taken. Third, the features and zones are designed around bulk properties, as described in Figure 7, which are less sensitive to printer to- printer variation. Finally, embedded firmware and system hardware were tuned during product development to generate subtle changes in media offset and shape relative to the emitter for each zone such that additional useful information may be extracted from the dataset.

This specific approach is summarized in Figure 7. For example, media in Zone 1 enters the sensor and obscures the Photo detector. Prior to Zone 1, the photo detector is saturated and the signal is low. When the leading edge of the media directly obscures the direct path between the emitter and the photo detector, a minimal amount of light is transmitted and the signal is high. As the media continues downstream, a larger area of the in-process media is exposed to the emitter and additional diffusely scattered light reaches the photo detector; the signal decreases. The output in Zone 1 is a strong function of media opacity and feed rate.

Further, media in Zone 3 is fed by two separate feed roller systems simultaneously. The relative velocity

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

of the roller systems is precisely controlled by embedded firmware to elicit a specific media response. The shape of the bubble is strongly coupled to a specific intrinsic property (basis weight). Heavier media are stiffer and are less likely to buckle; the upstream feed rollers will slip. Lighter media will buckle and the position of the sheet relative to the sensor will change. In this manner, the hardware and firmware within the system may be adjusted using expert domain knowledge to extract distinct information from the measurement based upon the dynamic response of the media to generated system inputs. This novel concurrent design approach allowed the photo detector to collect additional useful information that was strongly influenced by extrinsic system properties, Zk. Additionally, Zones 2, 3, and 4 extract similar information from the time series data. Each zone provides a distinct measure of media opacity that is a strong function of intrinsic media properties, 'k. This provides the algorithm with a degree of redundancy and robustness against gross error. Discretization of the analog data in this manner generated a richer feature set with some measurement redundancy. A small designed experiment was conducted to assess system performance and select the final feature set. Due to the information gained from the difference method previously described, inclusion of features from redundant zones yielded improved performance with minimal additional computing overhead. Features used for the machine learning algorithm are provided in Table I. Features x1; x2; : : : ; x5 are extrinsic system properties and uncontrollable external factors that are provided by the printer systems embedded firmware to help stratify and decouple the training set. Features x6; x7; x18 contain an abundance of useful intrinsic media information, but are nonlinearly coupled to Zk and Yi. These features are calculated from the raw data and contain minimum, maximum and mean calculations (a measure of opacity) and range calculations (a measure of uniformity). Features x19; x20; x21 and x22 represent the previously described difference calculations that are used to separate 'k information from the influence of Zk and Yi. This contention is supported by the different, distinct trends demonstrated by the plotted feature trends. However, all the features have significant boundary confusion, are not practical for use individually, but contribute to the overall classification performance.

V.IMPLEMENTATION PERFORMANCE

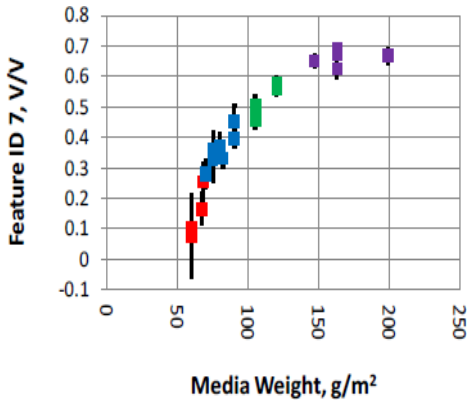
The results of the classification are given in Table II. The single node mean and the domain expert knowledge solutions are compared. The single node mean corresponds to the Zone 2 mean, or x8, and was selected as the best single node classification system. The domain expert knowledge system was compared against this implementation. In the case of the domain expert knowledge, a number of feature sets using different order kernels were evaluated in a designed experiment to select the optimum group. A second order polynomial kernel with the features shown in Table I was selected. The cost function of the algorithm was modified to ensure media near decision boundaries were classified in a manner that would have no negative impact on printer performance, as detailed in [23]. For this reason, “% Acceptable” is the key design metric for this system. This expert-prescribed cost function weighting Resulted in one particular paper type (Canon GFR-070) having poorer “% Correct” than in the single node mean case. This particular error is due to the fact that media is not naturally categorical. The weighting method was designed to integrate into the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

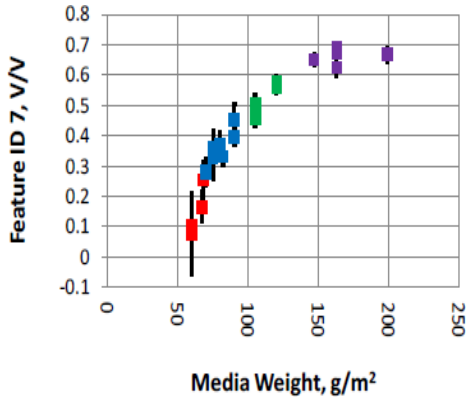
printers existing control scheme with minimal system impact. The richer feature set provides the machine learning algorithm more flexibility to adjust decision surfaces such that printer performance is not compromised when boundary confusion occurs.

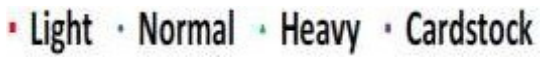
VI. CONCLUSIONS

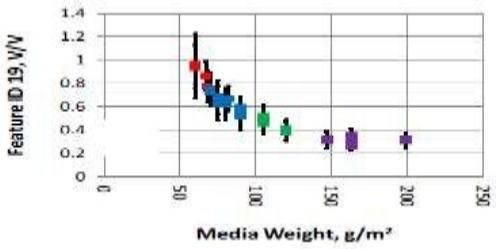
To further inform the design of Internet of Things (IoT) systems with domain expert knowledge and time series data, a

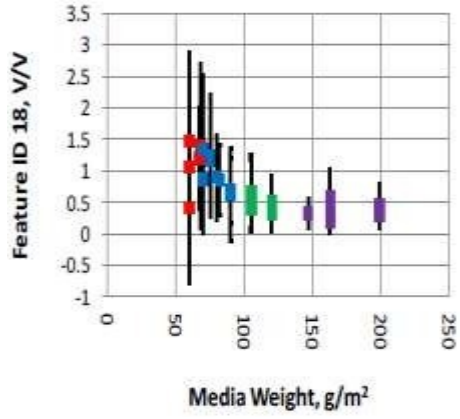


Methodology was developed.

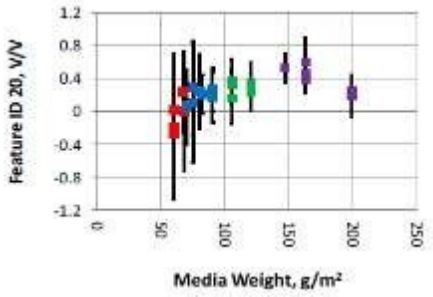








ISSN 2347–3657
Volume 9, Issue 4, Dec 2021



Scaled versions of some example input features across several media types are shown in Fig. 8. Although the features together include information for doing corner case separation, the features themselves suffer from boundary confusion (significantly overlapping error bars between categories). A system that is both reliable and accurate, but is also smaller, simpler, and cheaper than the alternatives. A mass-produced electro photographic printer served as an example of the methodology's application in a media classification system. When compared to a standard approach that did not use domain expert knowledge to enrich the dataset, the proposed methodology improved classifier accuracy by 16% and classifier acceptability by 6.5%. This approach can be employed by sensor-integrated Internet of Things (IoT) devices that want to take advantage of the performance gains afforded by modern sensor technology while also satisfying a number of market requirements.

REFERENCES

[1] N. Bajaj, G. T. C. Chiu, and J. P. Allebach, “Reduction of memory footprint and computation time for embedded support vector machine (SVM) by kernel expansion and consolidation,” in 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Sep. 2014, pp. 1–6.

[2] M. Brand, N. Oliver, and A. Pentland, “Coupled hidden Markov models for complex action recognition,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun
1997, pp. 994–999.

[3] T. Mori, Y. Nejigane, M. Shimosaka, Y. Segawa, T. Harada, and T. Sato, “Online recognition and segmentation for time-series motion with HMM and conceptual relation of actions,” in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Aug 2005, pp. 3864– 3870.

[4] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in Proceedings of the 2001 IEEE International Conference on Data Mining, 2001, pp. 289–296.

[5] E. Fuchs, T. Gruber, J. Nitschke, and B. Sick, “Online segmentation of time series based on polynomial least-squares approximations,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pp. 2232–2245, Sep. 2010.

[6] R. Agrawal, C. Faloutsos, and A. Swami, “Efficient similarity search in sequence databases,” in Foundations of Data Organization and Algorithms. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, Oct. 1993, pp. 69–84.

[7] K.-P. Chan and A. W.-C. Fu, “Efficient time series matching by wavelets,” in Proceedings of the 15th International Conference on Data Engineering, Mar 1999, pp. 126–133.

[8] I. Gu¨ler and E. D. U¨ beyli, “Multiclass support vector machines for EEGsignals classification,” IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 2, pp. 117–126, Mar. 2007.

[9] Y. Lei and M. J. Zuo, “Gear crack level identification based on weighted K nearest neighbor

ISSN 2347–3657
Volume 9, Issue 4, Dec 2021

classification algorithm," Mechanical Systems and Signal Processing, vol. 23, no. 5, pp. 1535–1547, Jul. 2009.

[10] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967.