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

***Mrs.T.Aswani*[1]*.,M.Ravali*[2]*.,M.Nichitha*[3]*.,B.Ruchitha*[4]*.,P.Madhavi*[5]**
1 Assistant Professor, Department Of Cse., Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India (aswani.thummalagunta@gmail.Com)
2, 3, 4, 5 B.Tech CSE, (19RG1A05F4, 19RG1A05F7, 20RG5A0513, 20RG5A0517),
Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

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 performanceand accuracy increase. Consequently, many electromechanicaldevices have incorporated sensor-enabled control schemes.Recently, machine Learning algorithms have begun to leveragethis trend to enable new functionality. Sensed information mayBe used to generate input features for algorithms that enableproactive diagnostics, system-awareness, and other morecomplex tasks such as classification. Concerns arise whenthe number of sensors and the capability of individual nodesare constrained due to cost or other associated factors likecomputation time and memory footprint. Previous efforts toaddress this concern have focused on a reduction of computationalrequirements during both the training and classification phases of embedded supervised machine learning algorithmdevelopment [1]. Methods attempting to minimize the numberof features required for classification also exist; these may beused to reduce the number of sensors necessary for a giventask.

This work presents a novel method to reduce the numberof sensors required for a supervised machine learning classificationsystem. Expert knowledge of expected sensor outputvariation as a function of intrinsic properties,

extrinsic properties,and uncontrollable external factors is used to establisha unique feature set that sufficiently decouples otherwise inseparableclasses. The system design and control system wereconcurrently tuned to elicit distinct dynamic responses withinpredefined temporal regions of a continuous data stream. Theanalog data was discredited into several distinct zones ofinterest corresponding to the sensors response to differentdynamical processes. A unique difference method allowedthe learning algorithm to extract additional useful information
From the confounded data set. This methodology is validatedby a case study of a print media classifier system developedfor a commercial laser printer, which was manufactured and deployed at a large volume. The resultant classification successexceeded that of embodiments using multiple sensors withonly a single sensor. Finally, the implications of this designmethodology and advantages over a traditional data-drivenclassification system are discussed.

II. BACKGROUND

The goal of simplification of multi-sensor systems by harvestingmore independent features from a reduced sensor setrelies on modification of the measured object usually basedon time or geometry. There are numerous studied methodsfor dimensionality reduction and representation of time seriesdata. General dimension-reduction and re-representationmethods include model-based techniques such as those usinghidden Markov models [2], [3]. A second class of methodshave attempted to reformulate the data with interpolative or regressionmethods such as piecewise linear (PLA) [4] or piecewisepolynomial (PPA) [5] approximations. Another groupof methods uses a symbolic representation optimized withcertain constraints such as symbolic aggregated approximation(SAX). Still other methods use transforms such as discreteFourier [6] or discrete cosine transforms or wavelet systems[7], [8]. Although these methods are largely designed foruse on general, potentially multi-dimensional time series, theyare frequently tested, presented, and verified on applicationspecificdata from medical data [8] to faults in mechanicalgear systems [9].Once the transformation has been performed, classificationtraining and evaluation can occur. Possible algorithms include1-nearest neighbors (1NN) or k-nearest neighbors (kNN) [10],which demonstrated considerable success when implementedwith representations like SAX in combination with dynamictime warping [11]. More sophisticated methods such as neuralnetworks, multi-layer perceptions [12], Bayesian networks[13], support vector machines [14] and decision trees [15] havealso been used with success and represent alternative designoptions. Some methods use information from a transformation,such as warping distance, as an additional feature and integratethis into the classification method [16]. In each case, thefeatures used to train these systems are selected to be as orthogonalas possible and the quality of the resulting algorithmis, amongst other things, a function of that orthogonality.Often, the system cannot be easily simplified, and hardwarewith embedded supervised machine learning systemsis designed using a complex network of various sensors. Intheory, this extra data enables the designer to build and test arobust algorithm since a network of sensors can be selected tomaximize feature orthogonality. This can lead to a temptationto deploy more sensors and computational resources than isstrictly necessary. In industries where customers are highlysensitive to product cost, such as office printing, the strategyis often to deploy a single sensor to partly meet design needs.These attempts have included using a set of electrodes totake electrical measurements of media [17], [18], a camerato measure surface roughness [19] or an ultrasonic sensor todetermine media density [20].

III. METHODOLOGY

In concurrently developed physical systems, the designerhas access to significantly more information about the situationthan is often available with analyzing time series data in ageneral case. Time series data output by a single sensor maycontain information about multiple physical quantities due tosystem dynamic behavior. Therefore, multiple physical quantitiesdo not always need to be measured by the same number ofphysical sensors. The designer has an opportunity to tune thehardware to produce a time series output from a single sensorand then discredited the output with domain expert knowledgeto produce multiple features while preserving orthogonality.

This results in a system with fewer sensor nodes and a lowerassociated 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 presentedset of features belongs to which one of several distinct sets).

The goal is to take as input a vector x 2 Rnf , where nf isthe number of features used for classification, and producean output y(x) which represents the classifier output. Givenxk 2 Rnf; k = 1; 2;:::; N are the feature vectors correspondingto N training examples and yk are the corresponding trueclasses (in this case yk = +1 if the measurement belongsto a set and yl = 1 if it does not), the classificationalgorithm is trained by solving the following optimizationproblem to determine a best separating hyper surfacedefined through a nonlinear mapping. Some interpretations of the LSSVMand other SVMs make assumptions about the variablesbeing independent and identically distributed random variables.While we cannot make this claim for this dataset dueto temporal correlation, SVM-type algorithms can still workwell 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, andit can be observationally inferred that an SVM might workwell given a geometric rather than probabilistic interpretationof 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 dualspace, the space of Lagrange multipliers of the constraints,_k (for k = 1; 2; N). b is a scalar bias offset term. Isa regularization parameter that can be used to control over fittingvs. under-fitting behavior, but was set as 1?W 2 Ruff isa vector of weights that, along with the mapping '(xk) helpsto define the decision hyper surface. The dual space classifier takes the form:

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

$K(x, x_k) = \varphi^T(x)\varphi(x_k)$ Is a Kernel function (a nonlinearmapping that allows additional flexibility in the classificationfunction). Both the dual space classifier and the solutionof the classifier optimization problem can be addressed byconsidering the Karsh-Kuhn-Tucker (KKT) conditions foroptimality:

$$\mathbf{w} = \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k),$$

$$\sum_{k=1}^{N} \alpha_k y_k = 0,$$

$$\alpha_k = \gamma e_k, \forall k = 1, 2, ..., N,$$

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

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

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

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 besolved for _ and b using standard methods (LU factorization,etc.). The Kernel function can take a number of different Forms, ofwhich $K(x, x_k) = x_k^T x$ (linear), $K(x, x_k) = (x_k^T x + \tau)^d$ (polynomial), and $K(x, x_k) = \exp\left(-\frac{||x - x_k||_2^2}{\sigma^2}\right)$

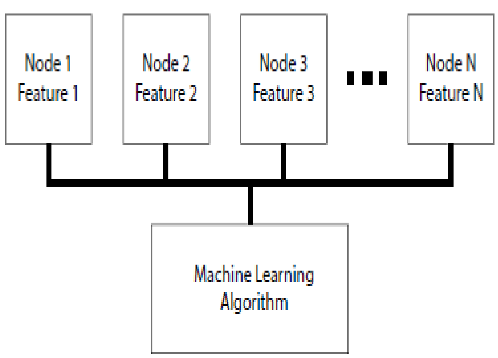


Fig. 1. Traditional method for enabling feature-based decision makingcapability on an existing device. The final classification algorithm is a functionof N features, represented by N nodes.In this work,only polynomial classifiers are considered. This is due to theapplication requirements of processing power and programmemory space, constrained to use the algorithm of [1].Typically, y(x) = +1 would yield a prediction that xbelongs in one set, and y(x) = □1 would correspond to theother complementary set. However, in some cases, includingmedia classification

in a printer, there are areas of the featurespace that for some comparisons make no difference (thereis 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 downstreamprocesses, and some that should be weighted more heavily.This idea was discussed and formulated into the training of amulticlass SVM problem and described in detail in [23]. Thesolution method for the system is the same. In this work, theresult associated with each classification is accordingly eitheran incorrect classification, an incorrect (but acceptable) error,or a correct classification. An acceptable error is simply onethat is tolerable to the downstream processes.In order to create a multiclass classification system, thedifferent classes are separated into complementary groups andevaluated in a one vs. all sense [22] (other options exist, butone vs. all is the encoding used in this work); if there are three

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

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

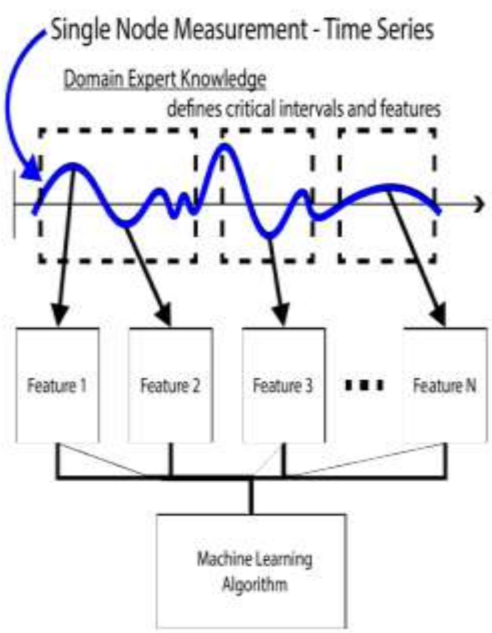


Fig. 2. The proposed approach uses the system knowledge of co-designed hardware to pull multiple features out of a single time series of data.

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

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] To2 Ropy).

Nape represents an ideal set of orthogonal intrinsic properties. _ _k. simply put, ideally, the sets to be classified are wellseparated by a measurement of some direct, relevant intrinsicphysical property and have good orthogonality. In the practicalcase, this is not so. Every measurement is a function ofboth the intrinsic property being measured and the propertiesof the physical system involved in that measurement. Theseproperties include the structure of the system and its operation,which are controllable by the system designer, and knownenvironmental factors which may not be controllable by thedesigner. Considering the form of the constructed intervals andcorresponding statistical measures, the training data examples ski are such that

$$
\begin{aligned}
x_k = [&f_1(\phi_k,Y_1,Z_k),\\
&f_2(\phi_k,Y_2,Z_k),\\
&\vdots\\
&f_N(\phi_k,Y_N,Z_k)]
\end{aligned}
$$

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 thatinfluence the measurement (Npe is the number of extrinsicproperties affecting measurements); and (Y1; Y2; YN),which are uncontrollable external factors that are a functionof the hardware design.

In the case of systems where measurements taken in differentintervals are coupled, taking the difference between Two functions can help to train the classifier with independentinformation about system interactions and decouple externalfactors that influence the measurement. This can be justifiedwith a brief expansion analysis. Given two functions if andfjord, the Taylor series expansions can be taken about a nominaloperating 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

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

For the same training example, __k = 0. The same is truefor _Ski. Therefore, the only remaining terms are those that Include _Yi and _I, the associated partial derivatives, andthe difference of the offset constants. This new feature, fi□fj ,is solely a function of _Yi and _Yj , which are functionsof certain fixed extrinsic system properties. This informationcan be learned by the classifier and improve classificationperformance.

IV. CASE STUDY AND IMPLEMENTATION

This case study applies the proposed approach to a commercialcolor 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 doadjust the media settings do so correctly. Incorrect settingson these devices may cause problems for both the customerand the manufacturer. To address this issue, an inexpensivesensor system and embedded machine learning algorithm wereimplemented to classify media without user input. The printercontrol system adjusted device parameters based on this mediaclassification.

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 streamcorresponding to the amount of light transmitted by in-processmedia. A simple model of the sensor was developed and,based upon this, system hardware and controls were tunedto generate an information-rich data stream by leveragingthe dynamic response of media to control system inputs.The printer generated features from this data stream foreach sheet of media. A broad population of standard officemedia with varied intrinsic properties, 'k, existing along acontinuum was sorted into 1 of 5 distinct classes: light,normal, heavy, card stock, and transparency. This dataset wasused to generate an embedded machine learning algorithm thatused these features to determine media class in near real time.Printer process parameters and system controls were adjustedbased upon this prediction. The final embodiment significantlyreduced overall cost, complexity, and system footprint whencompared to traditional implementations and is described ingreater detail in [24].A cross section of the printer media path is shown inFigure 3. The highlighted region contains a section view ofthe sensor and the surrounding printer hardware includingupstream feed rollers, media guides, and downstream feedrollers. The electrical design schematic for the optical sensorsystem is shown in Figure 4. Nominal circuit values weretuned to adjust the sensor gain, response, and sensitivity. Theresulting full scale range of the data set was maximized forthe population of expected media and maximum separation betweenmedia classes was achieved. Calibration was performedto compensate for system gain and offset errors.The sensor outputs a continuous data stream correspondingto the amount of infrared light transmitted by the in-processmedia. This output is a highly coupled function of manyconfounded factors including intrinsic media properties (e.g.,media basis weight, media roughness, media thickness, etc.)'K, extrinsic system properties (e.g., LED intensity, mediaspeed, media input source, phototransistor sensitivity, feedroller velocities, media shape and offset, LED and phototransistordirectionality, etc.) Zk and uncontrollable externalfactors (e.g., relative humidity, temperature, etc.) Yi. Figure5 depicts how variability caused by these confounded factorsimpacts the measurement for a single media. 20 measurementsfor a normal weight office media are shown. The signalvaries substantially from sheet to sheet and within a givensheet. Sensor output for a given media may vary as much as20% 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 muchas 60% of the sensors full scale range. This is primarily afunction of extrinsic system properties, Zk. Uncontrollableexternal factors, Yi, alter both the intrinsic media properties,

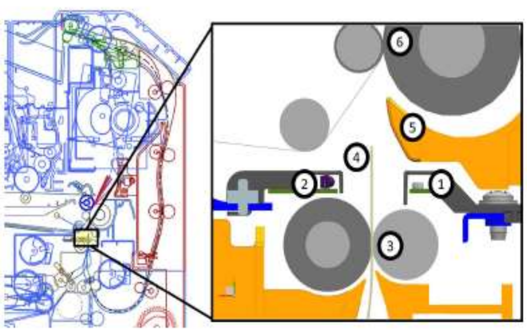


Fig. 3. A cross-section of the printer media path is depicted. The highlightedregion contains an optical sensor consisting of an LED (1) and phototransistor(2) that measures the amount of infrared light transmitted by a sheet of media(4) as it is processed by the printer. Media fed by upstream feed rollers(3) passes through the sensor, beyond a media guide (5), and into a set ofdownstream feed rollers (6). Hardware (physical design of the media path) And firmware (system timings and relative velocities of the feed rollers) weretuned during development to enhance data orthogonality by controlling theposition and shape of the media relative to the sensor in the spatial/temporal Domain.

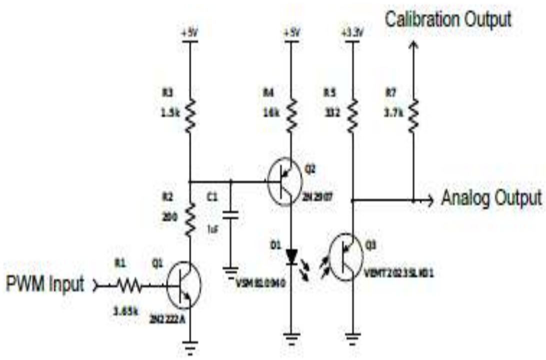


Fig. 4. The electrical design schematic for the optical sensor system is shown.The connection labeled "Analog Output" is the voltage signal measured bythe analog-to-digital converter and used in the classification system. Nominalcircuit values were selected to optimize the sensor gain, response, andsensitivity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

Figure 6 depicts how this variability manifests as boundaryconfusion. A broad set of standard office media possessing arange of intrinsic properties, 'k, existing along a continuumwere used to train and test the algorithm and are listed inTable II for reference. Corner cases (distinguishing light fromcard stock, for example) are easily distinguished. However,media properties exist along a continuum and variability fromsheet to sheet and within a given sheet made the classificationproblem particularly challenging. There was a large amount ofboundary confusion. This is especially true for the heavy classof media which significantly overlaps with both the normal and Card stock classes.

For a classifier to be successful, it must decouple therelevant intrinsic media properties, 'k, from the other confoundingvariables and generate a substantially orthogonalfeature set. Media to media variability must be decoupledfrom the variability seen from sheet to sheet or within a givensheet. For the case of media classification, this was achieved

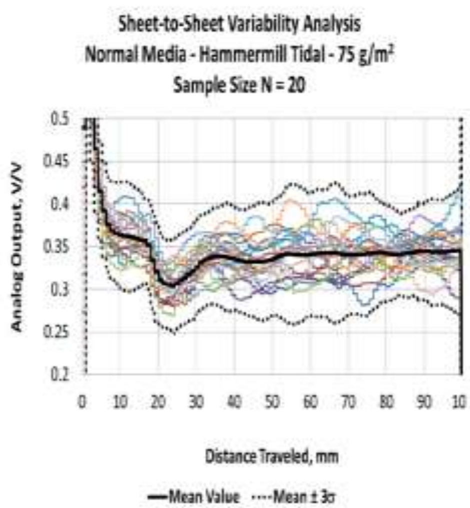


Fig. 5. Normalized analog sensor output for 20 separate sheets of a standardoffice paper are plotted. Data was collected for 100 millimeters of mediatravel. The population means and 99.7 percent confidence bands for this given Media are plotted for reference.

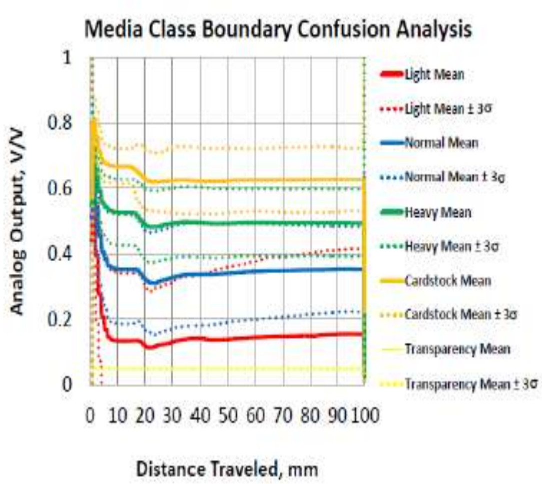


Fig. 6. Normalized analog sensor output for the mean and 99.7 percentconfidence band of each class are plotted. The population for each classconsists of 360 training samples from each media listed in Table II. A standardclassification problem utilizing a traditional feature set would be intractabledue to the continuous, overlapping nature of the data.By tuning system hardware and control parameters to leveragethe sensitivity of the measurement to uncontrollable externalfactors, Yi, and extrinsic system properties, Zk. Since thesensor output was a nonlinear function of 'k, Zk, and Yi,it was possible to use the dynamic response of the system tohelp decouple these convoluted variables using the differencemethod described previously.Concurrently developed printer control algorithms and sensorhardware were tuned during the development phase togenerate a continuous data stream that

could be deconstructedinto several distinct zones of interest corresponding to thesensors response to different dynamical processes. The resultanttime series data was divided into 5 distinct zones ofinterest that corresponded to changes in the printer processthat were designed to elicit a varied response from the sensor.In order to make the design more insensitive to printer-to printervariation, four ideas were considered when designingthe zone positions. First, a flag sensor (integrated into thepaper feed control system) allowed accurate registration ofthe leading edge of the sheet, and the traverse distance wasknown from the paper feed drive encoders. Second, the zonesare larger than strictly necessary for a single printer in orderto accommodate variation around the population of printers(determined empirically from a number of different printers).

It is important to be aware that performance can decreaseif the buffer regions are too large as the data quality willdecrease from the statistical measure being taken. Third, thefeatures and zones are designed around bulk properties, asdescribed in Figure 7, which are less sensitive to printer to-printer variation. Finally, embedded firmware and systemhardware were tuned during product development to generatesubtle changes in media offset and shape relative to the emitterfor each zone such that additional useful information may beextracted from the dataset.

This specific approach is summarized in Figure 7. Forexample, media in Zone 1 enters the sensor and obscures thePhotodetector. Prior to Zone 1, the photodetector is saturatedand the signal is low. When the leading edge of the mediadirectly obscures the direct path between the emitter and thephotodetector, a minimal amount of light is transmitted and thesignal is high. As the media continues downstream, a largerarea of the in-process media is exposed to the emitter andadditional diffusely scattered light reaches the photo detector; thesignal decreases. The output in Zone 1 is a strong function ofmedia opacity and feed rate.

Further, media in Zone 3 is fed by two separate feedroller systems simultaneously. The relative velocity of theroller systems is precisely controlled by embedded firmwareto elicit a specific media response. The shape of the bubble isstrongly coupled to a specific intrinsic property (basis weight).Heavier media are stiffer and are less likely to buckle; theupstream feed rollers will slip. Lighter media will buckle andthe position of the sheet relative to the sensor will change.In this manner, the hardware and firmware within the systemmay be adjusted using expert domain knowledge to extractdistinct information from the measurement based upon thedynamic response of the media to generated system inputs.This novel concurrent design approach allowed the photodetectorto collect additional useful information that was stronglyinfluenced by extrinsic system properties, Zk.Additionally, Zones 2, 3, and 4 extract similar informationfrom the time series data. Each zone provides a distinctmeasure of media opacity that is a strong function of intrinsicmedia properties, 'k. This provides the algorithm with adegree of redundancy and robustness against gross error.Discretization of the analog data in this manner generateda richer feature set with some measurement redundancy. Asmall designed experiment was conducted to assess systemperformance and select the final feature set. Due to theinformation gained from the difference method previouslydescribed, inclusion of features from redundant zones yieldedimproved performance with minimal additional computingoverhead. Features used for the machine learning algorithmare provided in Table I. Features x1; x2; : : : ; x5 are extrinsicsystem properties and uncontrollable external factors that areprovided by the printer systems embedded firmware to helpstratify and decouple the training set. Features x6; x7; x18contain an abundance of useful intrinsic media information,but are nonlinearly coupled to Zk and Yi. These features arecalculated from the raw data and contain minimum, maximumand mean calculations (a measure of opacity) and range calculations(a measure of uniformity). Features x19; x20; x21 andx22 represent the previously described difference calculationsthat are used to separate 'k information from the influence ofZk and Yi.This contention is supported by thedifferent, distinct trends demonstrated by the plotted featuretrends. However, all the features have significant boundaryconfusion, are not practical for use individually, but contributeto the overall classification performance.

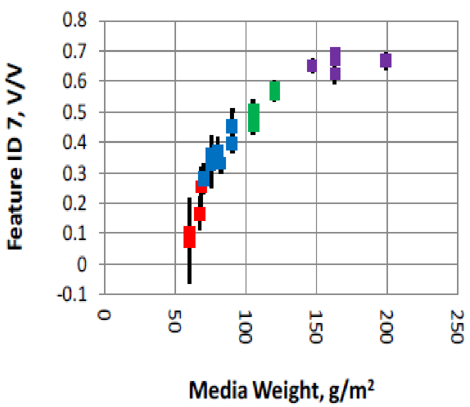
V.IMPLEMENTATION PERFORMANCE

The results of the classification are given in Table II. Thesingle node mean and the domain expert knowledge solutionsare compared. The single node mean corresponds to the Zone2 mean, or x8, and was selected as the best single nodeclassification system. The domain expert knowledge systemwas compared against this implementation. In the case of thedomain expert knowledge, a number of feature sets using differentorder kernels were evaluated in a designed experiment toselect the optimum group. A second order polynomial kernelwith the features shown in Table I was selected. The costfunction of the algorithm was modified to ensure media near decision boundaries were classified in a manner that wouldhave no negative impact on printer performance, as detailed in[23]. For this reason, "% Acceptable" is the key design metricfor this system. This expert-prescribed cost function weighting

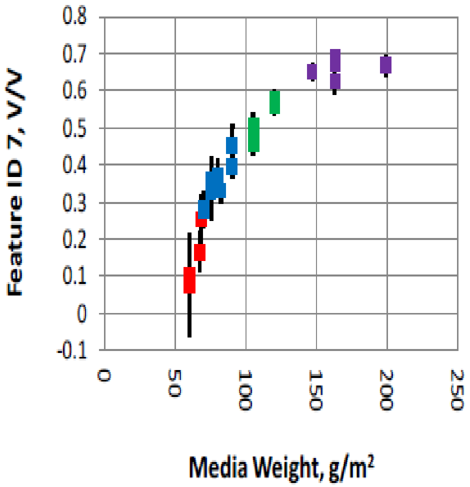
Resulted in one particular paper type (Canon GFR-070) havingpoorer "% Correct" than in the single node mean case. Thisparticular error is due to the fact that media is not naturallycategorical. The weighting method was designed to integrateinto the printers existing control scheme with minimal systemimpact. The richer feature set provides the machine learningalgorithm more flexibility to adjust decision surfaces suchthat printer performance is not compromised when boundaryconfusion occurs.

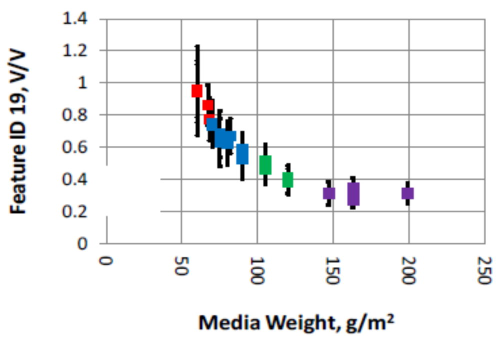
VI. CONCLUSIONS

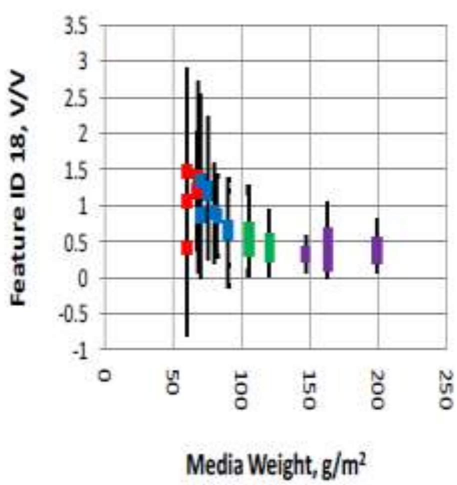
To further inform the design of Internet of Things (IoT) systems with domain expert knowledge and time series data, a

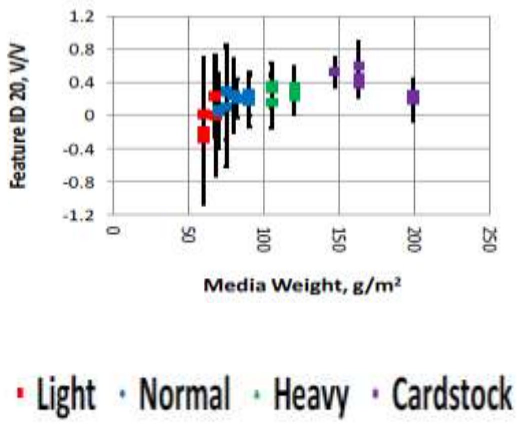


Methodology was developed.


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
Feature ID 7, V/V
0
50
100
150
200
250
Media Weight, g/m²


1.4
1.2
1
0.8
0.6
0.4
0.2
0
Feature ID 19, V/V
0
50
100
150
200
250
Media Weight, g/m²


3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
Feature ID 18, V/V
0
50
100
150
200
250
Media Weight, g/m²



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 memoryfootprint and computation time for embedded support vector machine(SVM) by kernel expansion and consolidation," in 2014 IEEE InternationalWorkshop on Machine Learning for Signal Processing (MLSP),Sep. 2014, pp. 1–6.
[2] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov modelsfor complex action recognition," in Proceedings of the IEEE ComputerSociety 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 HMMand conceptual relation of actions," in 2005 IEEE/RSJ InternationalConference on Intelligent Robots and Systems, Aug 2005, pp. 3864–3870.
[4] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm forsegmenting time series," in Proceedings of the 2001 IEEE InternationalConference on Data Mining, 2001, pp. 289–296.
[5] E. Fuchs, T. Gruber, J. Nitschke, and B. Sick, "Online segmentation oftime series based on polynomial least-squares approximations," IEEETransactions 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 searchin sequence databases," in Foundations of Data Organization andAlgorithms. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg,Oct. 1993, pp. 69–84.
[7] K.-P. Chan and A. W.-C. Fu, "Efficient time series matching bywavelets," in Proceedings of the 15th International Conference on DataEngineering, Mar 1999, pp. 126–133.
[8] I. Gu¨ler and E. D. U¨ beyli, "Multiclass support vector machines for EEGsignalsclassification," IEEE Transactions on Information Technology inBiomedicine, vol. 11, no. 2, pp. 117–126, Mar. 2007.
[9] Y. Lei and M. J. Zuo, "Gear crack level identification based on weightedK nearest neighbor classification algorithm," Mechanical Systems andSignal 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.

