# Adaptation of smard grid technologies

**Conference Paper** *in* IEEE International Conference on Fuzzy Systems · June 2012
DOI: 10.1109/FUZZ-IEEE.2012.6251193



**4 authors**, including:

Sergey Petrovski
Samara State Technical University


Andrei Petrovski
Robert Gordon University

# Intelligent Measurement in Unmanned Aerial Cyber Physical Systems for Traffic Surveillance

Andrei Petrovski[1(✉)], Prapa Rattadilok[1], and Sergey Petrovskii[2]

[1] School of Computing Sciences and Digital Media, The Robert Gordon University, Aberdeen, UK
{a.petrovski,p.rattadilok}@rgu.ac.uk
[2] School of Electric Stations, Samara State Technical University, Samara, Russian Federation
petrovski@rambler.ru

**Abstract.** An adaptive framework for building intelligent measurement systems has been proposed in the paper and tested on simulated traffic surveillance data. The use of the framework enables making intelligent decisions related to the presence of anomalies in the surveillance data with the help of statistical analysis, computational intelligent and machine learning. Computational intelligence can also be effectively utilised for identifying the main contributing features in detecting anomalous data points within the surveillance data. The experimental results have demonstrated that a reasonable performance is achieved in terms of inferential accuracy and data processing speed.

**Keywords:** Intelligent measurement · Traffic surveillance · Data anomalies · Computational intelligence · Artificial neural networks · Cyber physical system

## 1 Introduction

One of the main purposes of intelligent measurement systems (IMS) is to model the relationship between information that is required ('primary characteristics'), and the information which may be readily derived from (processed) sensor outputs such as target tracks ('secondary variable'). An IMS is capable of providing frequent 'on-line' estimates of primary characteristics on the basis of their correlation with the data, obtained from available sensors, measured in real time. As such, an IMS can help to reduce the need for measuring devices, improve system reliability, and develop tight control policies.

There are several advantages of IMS in comparison with traditional instrumentation [3]:

- Such measurement systems give more insight into the process under observation through capturing the information hidden in data.
- They are an emergent technology that allows users to improve productivity, become more energy and cost efficient.


C. Jayne and L. Iliadis (Eds.): EANN 2016, CCIS 629, pp. 161–175, 2016.
DOI: 10.1007/978-3-319-44188-7_12

- They can be easily implemented on existing hardware; moreover, various model-building algorithms can be used to adapt the IMS when an operating environment changes.
- They involve little or no capital cost such as the cost of installation, management of the required infrastructure, and commissioning.

The range of tasks fulfilled by IMS is quite broad – not only can IMS be used as a substitute or complement to physical sensors, but they can also perform monitoring and control of the process under observation, and can provide off-line operational assistance (e.g. design, diagnosis, knowledge refinement) [2].

The key challenge in building an IMS is to find a suitable structure for the inference model(s), using which a good estimator of the primary characteristics could be found. A basic rule in estimation is not to estimate what is already known or can be inferred from the data available. In other words, it is important to be able to utilise prior knowledge and physical insights about the process under observation/analysis when selecting the model structure. It is customary to distinguish between three levels of prior knowledge [4]:

- *White-box* models: the structure and parameters of the model are known or can be obtained from physical insights or basic principles;
- *Grey-box* models: some physical insights are available, but several model parameters remain to be determined from observed data;
- *Black-box* models: no physical insight is available, but the chosen model structure belongs to generic classes (e.g. artificial neural networks) that are known to have good flexibility and have been successfully applied in various problem domains.

Most of the existing IMS utilise black-box models operate on sensor data and produce estimates of essential (or primary) characteristics of the system under observation – for example, an unmanned aerial system (UAS) as shown in Fig. 1. (N.B. The

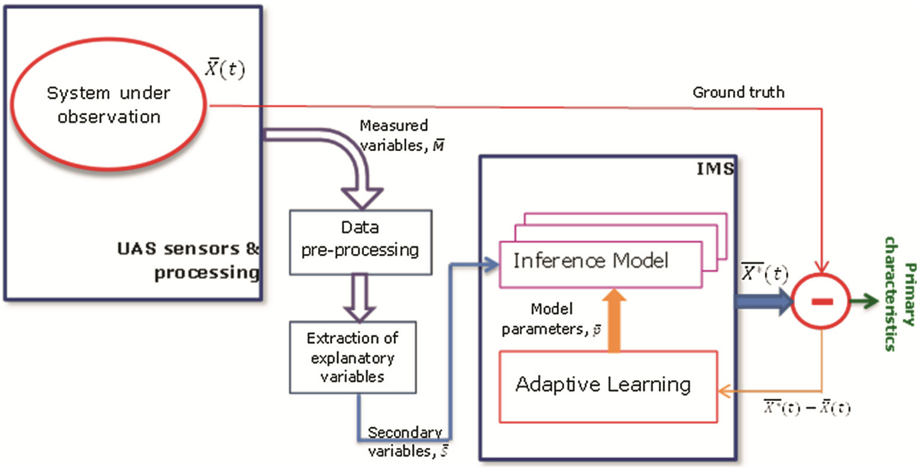


**Fig. 1.** IMS framework for the unmanned aerial vehicle

red arrow on the diagram representing ground truth information is desirable for more effective leaning, but not mandatory for the operation of an IMS.)

Having determined the relationship between the primary characteristics and the secondary variables, it becomes possible to obtain reasonable estimates of the former much faster and at a lower cost.

Also, the ability to infer primary characteristics raises the level of "information intelligence" coming from the UAS, enabling thereby to shift the workload of ground operators from "target detection to target analysis" and to optimise the throughput of data communication channels. Taking a road traffic example where it is desired to identify dangerous drivers represented by the state vector $\overline{\mathrm{X}}(\mathrm{t})$, the 'dangerous driver' categorisation would be the primary characteristic, while secondary variables could include such quantities as driving speed or lane discipline.

A UAS in this context can be considered an autonomous cyber physical system that is used to acquire large amount of data about complex and changing environments, to perform interpretation and fusion of the data, and to present the information gathered or inferred in a synthetic and compact form highlighting the features of interest in the environment explored. The situation awareness of a UAS is determined by its operating conditions, various inputs obtained from essential sensors, as well as control adjustments received from a ground station. The situation awareness in terms of determining abnormal traffic conditions is an example of a primary characteristic that is difficult to measure directly. However, the large amount of data coming from on-board sensors or received from a ground station can be referred to as secondary variables. Due to the nature of UAS operation, the states of many secondary variables reflect the states of primary characteristics. For instance, surveillance data obtained from various sensors can indicate, and even identify, unusual or dangerous behaviour of drivers on the road [6].

Heterogeneous data acquiring sensors on-board of an unmanned aerial vehicle, which is part of the UAS, also add complexity in the form of analytical challenges, especially when there exist time and cost differences in processing data from different sources. Selecting suitable data acquisition sources, e.g. data that can be processed approximately in order to obtain representative samples, can help in time critical situations. Additional data acquisition sources that involve longer data processing but are more accurate or detailed, can be applied later to provide adaptive measurement features.

With the vast amounts of data, traditional data acquisition and data processing methods have become inefficient or sometimes inappropriate, especially in a real time environment. Computational Intelligence (CI) techniques have been successfully applied to problems in various application domains [1, 5]. These techniques however require accurately labelled training data to provide reliable and accurate specification of the context in which a UAS operates. For example, drivers may behave differently in the different road conditions (e.g. icy, wet, and foggy). The term "driver(s)" used throughout this paper refers to drivers of vehicles on the road (i.e. in the simulated model) under the surveillance of a UAS. The context enables the system to highlight potential anomalies in the data so that intelligent and autonomous control of the underlying process can be carried out.

Anomalies are defined as incidences or occurrences, under a given circumstances or a set of assumptions, that are different from the expectance. By their nature, these incidences are rare and often not known in advance. This makes it difficult for the computational intelligence techniques to form an appropriate training dataset. Moreover, UAVs often operate in different or dynamic environments [11]. This can further aggravate the lack of training data by the increased likelihood of intermittent anomalies. Computational intelligence techniques that are used to tackle dynamic problems should therefore be able to adapt to environmental/contextual changes [6].

The research work described in the presented paper is aimed at using machine learning algorithms for addressing 'Situational Assessment'. The immediate application area is the development and evaluation of such algorithms for a UAS application carrying out wide area surveillance of a tract of ground.

The detection of unusual profiles or anomalous behavioural characteristics from sensor data is especially complicated in security applications where the threat indicators may not be known in advance [8]. Data-driven modelling in such cases can yield insights on usual and baseline profiles, which in turn can be used to isolate unusual profiles when new data are observed in real time.

In general terms, therefore, the problem being tackled can be defined as finding the most effective ways of using measured data obtained from multiple sensors on board an aerial vehicle, in order to address the inherent difficulty in precisely defining and quantifying what constitutes anomalies. The presence of several sources of variability in anomalous patterns (for example, traffic density, vehicle types, features of terrain, etc.) and the limited availability, or even absence, of training datasets aggravate the difficulty of the problem being addressed [10].

The desired outcome of the work would be to devise a solution framework for intelligent processing of data obtained from multiple UAS sensors. This framework, described in Sect. 2 of the paper, is built with the premise that all the data sources considered together are capable of capturing the important features that could lead to a reliable anomaly detection, to efficient extraction and to intelligent interpretation of these features, which could in turn significantly reduce the number of false alarms generated as a result of the UAS operation.

To handle the challenges presented by the problem being addressed an incremental approach was adopted as a three-stage development of intelligent measurement systems. The first stage (anomaly detection) processes the available data by extracting the most representative features (referred to as 'secondary' variables) that characterise potential anomalies – this process is described in Sect. 3. For anomaly detection a mixture of statistical analysis and computational intelligence (CI) techniques has been adopted. The choice of detection techniques depends on the amount of historical data and the availability of insights on 'normal' system profiles – at the start of the detection process preferences are given to statistical techniques utilising probabilistic measure of data anomalies. As more data is being obtained, anomalous patterns/profiles start appearing, which can be detected more effectively with the help of CI techniques.

The features selected are then used to build inferential models, demonstrated in Sect. 4, that are utilised in the second stage (anomaly modelling) to interpret the new incoming data for real-time decision. In the second stage of building data-driven inference models two types of classifiers have been used – conventional classifiers utilising clustering algorithms, which do not require training data sets, and computational intelligence methods that carry out supervised learning of anomalous data patterns (in particular, artificial neural network (ANN)).

Finally, when the operating conditions of the system/process under observation change, both the secondary variables and the inference models are adapted in the third stage (anomaly modification) to provide the means of adjusting the IMS within dynamic operating environments. This final stage of adaptive measurement by the IMS is implemented using an automated machine learning algorithm, described in our previous work [7], that continuously tunes the inference models built for processing measured data and the representative features of data anomalies.

The proposed approach to intelligent measurement is evaluated on simulated and benchmark datasets – the main conclusions and proposed areas of further research are summarized in Sect. 5.

## 2 Inferential Measurement Systems

The impediments caused by unavailability or ineffectiveness of conventional measurements can negatively affect "situational assessment", but the problem can be alleviated, at least partially, by developing an intelligent measurement system (IMS) that performs intelligent sensing through the use of "soft" sensor technology. Intelligent sensing is a relatively new capability of measurement systems that supports such features as long mission duration, reliability and availability, real-time operation in hazardous and changing environments, as well as flexibility of use. These requirements lead to measurement systems with increasingly autonomous functionalities based on decentralised and distributed system architecture, effectively utilising available instrumentation data. Figure 2 illustrates a generic framework for building an IMS, proposed in [7].

Modelling using Computational Intelligence (CI) has become a versatile tool for enhancing the capabilities and efficiency of inferential measurement systems [5]. This type of modelling utilises the computational capabilities of modern computing devices (smart sensors, DSP-based microcontroller, and microprocessors) to effectively process the acquired input and infer the desired information. The AI-based techniques are applicable at various layers of IMS – from the data acquisition (sensor) layer, through to the layer of instrument calibration and customisation, then to the layer of process modelling, control and optimisation, and finally to the knowledge acquisition layer [6]. The wide spectrum of possible applications is due to the capabilities of an IMS to gain insight into the behaviour of complex dynamic systems by means of data-driven modelling, a systematic approach to which is described in this section.

The underlying principle of "soft" sensing is in estimating unmeasured variables, properties or parameters by using a model of a process under investigation, or of a part thereof, that correlates the measurements of interest (primary characteristics in Fig. 1) with more immediate (secondary) variables. As the name suggests, the model used by "soft" sensors is usually implemented in software; the secondary variables for such

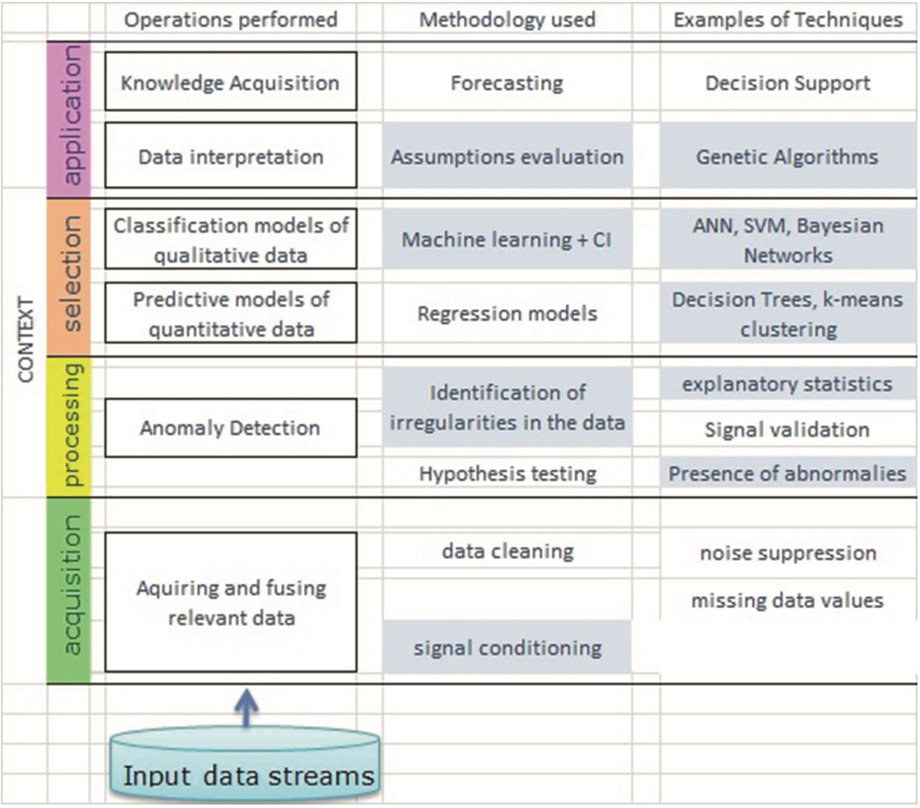


**Fig. 2.** Generic framework for building Intelligent Measurement Systems (IMS)

sensors are the controlled inputs, disturbances, and other intermediate variables affecting the process/application of interest [9].

In the course of the project the following tasks have been addressed:

- *Data pre-processing*: this step is performed only for building the inference models based on supervised machine learning techniques. Essentially, the step involves annotating the input data streams with the "ground truth" values needed for training certain Computational Intelligence (CI) algorithms (discussed later in the paper). Once the necessary training has been carried out, the data pre-processing activity becomes unnecessary, but can still be used if the reduction of noise in the data streams or filling in missing values are desired.
- *Selection of secondary* variables: it is important to choose the appropriate secondary variables (also referred to as data filters) to be used in building the inference model(s) – the number of these variables affects the time and complexity of inference, as well as the size of the data set needed for model development. The main objective of this step is to make use of the least number of secondary variables to develop a model of sufficient accuracy.

- *Building inference models*: Once a set of potential secondary variables is selected and their values are determined (this might involve passing the original data streams coming from sensors through several data filters), inferential models can be obtained using various data-driven modelling paradigms. At this stage it is important to strike the right balance between the accuracy and generalizability (i.e. minimising the effect of overtraining), and simplicity of the inference models. This is often achieved by varying the number of secondary variables (e.g. number of input nodes of ANNs) used in building the models through running screening and regression experiments (explained in more detail later in the paper).
- *Evaluating and tuning the inference models*: the inference models built have been validated on previously unseen data using a cross-validation approach. After the validation the inference model parameters (e.g. the window size of a data filter) can be dynamically adjusted if the operating environment changes (e.g. significant increase in traffic density) or the objectives of inferential modelling are modified (e.g. switching from the identification to classification mode of operation). The process of dynamic parameter adjustment is shown by the block at the bottom in Fig. 1, and is performed by a meta-learning layer of the developed IMS using a genetic algorithm (one of the Computational Intelligence techniques adopted within the proposed framework) [7].

### 2.1 Context Acquisition Level

In the presented research work, it is assumed that raw input data are pre-processed by having been already passed through the stages at the Data Acquisition level in Fig. 2 (e.g. data cleaning, fusing) and therefore this level is not considered. The only exception is the data discretisation activity, which can also be attributed to context processing level.

### 2.2 Context Processing Level

The Context Processing level in Fig. 2 utilises statistical and mathematical techniques of characterising raw input data. Depending on the complexity of the application domain, statistical methods can be used with the raw input data in order to identify anomalies within the input data stream; alternatively, statistical analysis may be used to prepare the raw input data for processing by computational intelligence techniques in identifying the pattern(s) of interest (or anomalies).

At this level, measurable variables are used to create secondary variables by applying different data filters and window sizes. For example, a secondary variable of *speed* may be defined as the change in distance travelled over *a* period of time, where *change* represents a data filter, *period of time* represents a window size ($W_1 = 10$ *sec*) applied to the measurable variable *distance*. Secondary variables can also be obtained by nesting data filters (with corresponding window sizes) one within another. For example, a composite secondary variable, based on the one exemplified above, could be defined as an average over the observed length of the road of the changes in travelled distance in a specified period of time. The applications of data filters and window sizes onto measurable variables are carried out by the Context Processing (see Fig. 2). Context agnostic data filters can also be created that characterise interactions between objects within the system under observation (e.g. relative distances or speeds) or the operating environment (e.g. object density).

Context Processing might also involve data annotation, which provides ground truth for the training of supervised learning techniques and for evaluating the accuracy of both supervised and unsupervised learning. Ground truth labels can be obtained by using some form of statistical thresholds (e.g. $3\sigma$ interval for normally distributed data), by manual annotation, or by obtaining the labels directly from a simulation model.

### 2.3 Context Selection Level

Once the data anomalies have been identified, they are then passed onto the Context Selection level. Classification of anomalies and the predictions of their effects are achieved by applying machine learning in order to build inference models. Additional raw or processed input data may be required at this level.

The Inference Model builder operates in the following way:

- The structure of the model specifies which learning technique $\mathcal{L}$ is going to be used with the chosen secondary variables $S_j, j \subset \overline{1, m}$.
- The specified learning technique checks the need for data conditioning and training datasets.
- The selected secondary variables determine the measurable variables $M_i, i \subset \overline{1, n}$ and the data filters with corresponding window sizes $f^k_{w_k}, k \subset \overline{1, l}$.
- This process minimises the amount of data collected and processed while the inference models (represented as tuples $(\mathcal{L}, \overline{\mathrm{S}}, \overline{\mathrm{p}})$, where $\overline{\mathrm{S}}$ is the vector of secondary variables and $\overline{\mathrm{p}}$ is a vector of parameters for the learning technique $\mathcal{L}$ (for example as an error acceptance rate for ANN) being built and evaluated.

The number of selected secondary variables $m$ directly influences the structure, complexity and usability of the inference model, and thus needs to be optimized in accordance with the size of data samples.

### 2.4 Context Application Level

The Context Application level supports autonomous operation of the IMS by reducing the importance of human involvement in adjusting the model to changing operating conditions. As was mentioned previously, this task is achieved with the help of genetic algorithms, which autonomously select the optimal parameters on the inference through the effective use of evolutionary processes adopted from nature.

Based on the way the intelligence is obtained, intelligent measurement systems can be categorised either by the function they perform (calibration, error compensation, data validation, anomaly detection, adaptation, decision making, etc.) or by the technique(s) used (statistical, symbolic, ANN-based, fuzzy logic, and the like) [2]. Having chosen the secondary measurands to be used, the processed data together with the inference models build are then passed on to an autonomously chosen supervised or unsupervised learning algorithm. These learning algorithms are used to identify and classify the patterns of interest in the analysed data streams, which reflect dynamic operating environment.

## 3 Data Filtering for Intelligent Instrumentation

The analysis of surveillance information in general, especially related to situation awareness, is a complex process that, given the amount and heterogeneous nature of data, is prone to data overload. This results in an inability to support real-time processing and analysis of surveillance data. This is especially true when using mobile platforms where datalink and bandwidth issues are significant [12, 13].

### 3.1 Problem Specification

In order to design and build an intelligent measurement system a testing dataset derived from a MATLAB vehicle simulation model (developed and evaluated by our industrial collaborator) was used in this research. This model is capable of mimicking the behaviour of various types of drivers; typical examples are the normal and "cowboy" drivers. Normal drivers are those that observe road discipline, which regulates that no undertaking is acceptable, and that the vehicles shall move to the left lane whenever possible. The "cowboy" drivers are those that might violate these constraints.

The simulation model provides ground truth 'normal' and 'cowboy' labels; the characteristics of particular drivers within a type are subject to distributions rather than being entirely deterministic – frequencies and instances of exhibited behaviours are context dependent (e.g. traffic density, behaviours of other close vehicles). Therefore, a "cowboy" driver may or may not exhibit the salient features of his behaviour during the observation period.

In total, five driver types are considered – three of these are additional 'abnormal' types (viz. slow, cautious and boy racer). The slow and cautious drivers are similar to the normal driver in that they both follow the lane discipline. Cautious drivers, however, tend to leave a larger gap in front of them, whereas the slow drivers move more slowly, as well as react, brake and accelerate more gently. The "boy racers" are similar to the "cowboy" drivers in that both types do not always follow the lane discipline; what distinguishes them is that the "boy racers" drive faster, braking and accelerating harder, than the "cowboy" drivers.

Such a vehicle simulation model creates a data source rich enough to be used for making intelligent measurement of the driver type. In particular, the presence of several types of anomalous drivers makes it sensible to conduct the inference process in a number of phases: identification, classification and prediction. The identification phase minimises the volume of data and the data processing cost by analysing only a small set of measured data using anomaly identification techniques, such, for instance, as outlier detection. Identified potential anomalies are then passed onto the classification phase, where they are separated out into different types.

As a means of understanding the potential of the IMS techniques developed in the general context, the aims of such evaluation are to use the datasets generated by this vehicle simulation model in order to:

1. Identify anomalous drivers (i.e. all driver types different from the "normal") – the identification phase of IMS operation.
2. Appropriately classify these anomalous drivers into the corresponding types.

### 3.2 Choosing Secondary Measurands

There are a number of simulation parameters that can be adjusted within the MATLAB traffic simulation model. Some of the simulation parameters directly affect the behaviour of simulated drivers (i.e. *speed ranges*, driver *reaction time*). The other parameters determine the environment – in our case the characteristics of the road (i.e. *lane width* and *number of lanes*), which indirectly influence how each driver behaves.

The task of choosing the right set of variable to measure (i.e. measurands), which provide reliable inference capabilities, is not trivial. Therefore, selection of an appropriate set of secondary (i.e. based on applying filtering to directly measurable data inputs) measurands is a vital step in building an inference measurement system, affecting its accuracy, complexity and generalizability of the inference operation(s).

A conventional methodology of choosing a set of input variables is based on conducting a 'screening' experiment aimed at establishing the significance of each input in terms of inference capabilities of an IMS. This experiment is done by setting the high and low levels for six main variables within the vehicle simulation model: *lane*, *average speed*, *traffic rate*, *road length*, *road width* and *reaction time*. The proportion of normal vs. anomalous drivers was fixed as 80:20. There are sixteen trials in total, i.e. half-factorial screening experiment has been carried out.

Given the difficulty of empirically selecting secondary measurands for building an inference model(s), a more systematic approach has been proposed in the course of this work that is capable of not only choosing the most appropriate input data streams and associated data filters, but also of automatically determining the most effective learning algorithms for adapting the IMS to operate in changing environments. The results of the screening experiments are summarised in Table 1.

The results in Table 1 are obtained using four different statistical data filters (i.e. AVERAGE, VARIANCE, MIN, MAX) on three measurable variables (*distance travelled (along road)*, *lateral movement* or frequency of changing lanes, and *total number* of vehicles). The results shown are obtained using balanced training datasets, which use the equal number of training examples for each driver type (unbalanced training datasets use unequal number of training examples).

The F-test and t-test have been applied to analyse the statistical significance of different features, represented by low p-values (which represent the probability of obtaining the observed differences in accuracy purely by inherent randomness of experiments). Low p-values $p \leq 0.05$ indicate that the differences in model performance are

**Table 1.** Significance level of secondary measurands

| Input data | p-values | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Neural Networks | | SVM | | Bayesian networks | | K-Means | |
| | Balan. | Unbal | Balan. | Unbal | Balan. | Unbal | Balan. | Unbal |
| lane | 0.012 | 0.216 | 0.006 | 0.088 | 0.001 | 0.132 | 0.036 | 0.105 |
| average speed | 0.053 | 0.388 | 0.076 | 0.193 | 0.013 | 0.219 | 0.214 | 0.476 |
| traffic rate | 0.321 | 0.916 | 0.346 | 0.168 | 0.194 | 0.341 | 0.912 | 0.171 |
| road length | 0.132 | 0.439 | 0.012 | 0.196 | 0.002 | 0.228 | 0.172 | 0.313 |
| road width | 0.021 | 0.437 | 0.035 | 0.251 | 0.002 | 0.322 | 0.256 | 0.195 |
| reaction time | 0.588 | 0.847 | 0.121 | 0.777 | 0.588 | 0.346 | 0.269 | 0.176 |
| lane*avg speed | 0.097 | 0.421 | 0.032 | 0.354 | 0.002 | 0.644 | 0.205 | 0.569 |
| lane*traffic rate | 0.020 | 0.415 | 0.013 | 0.152 | 0.002 | 0.126 | 0.107 | 0.486 |
| lane*road length | 0.083 | 0.338 | 0.100 | 0.200 | 0.004 | 0.241 | 0.171 | 0.234 |
| lane*road width | 0.022 | 0.616 | 0.017 | 0.796 | 0.008 | 0.869 | 0.150 | 0.694 |
| lane*reaction time | 0.104 | 0.759 | 0.283 | 0.518 | 0.025 | 0.829 | 0.913 | 0.236 |
| avg speed*road length | 0.468 | 0.488 | 0.186 | 0.842 | 0.003 | 0.864 | 0.079 | 0.927 |
| avg speed*reac.time | 0.062 | 0.573 | 0.028 | 0.998 | 0.003 | 0.646 | 0.845 | 0.821 |

attributed to systematic factors (significant parameters are highlighted in yellow, low p-values are shown in red):

- *Lane* is the only variable in this experiment that is shown to significantly affect the accuracy measure of all learning techniques.
- Another variable that has a significant effect on the accuracy rate of supervised learning techniques is the *road width*;
- The interactions between *lane* with *road width* and *traffic rate* significantly affect the effectiveness of supervised learning – see the p-values highlighted in red in the table below.

These significance values exhibit a degree of correlation with the design of the vehicle simulation model, where lane discipline is a major characteristic that distinguishes different types of drivers.

## 4 Building Inferential Capabilities Within IMS

The analysis of surveillance information in general, especially related to situation awareness, is a complex process that, given the amount and heterogeneous nature of data, is prone to data overload. This results in an inability to support real-time processing and analysis of surveillance data. This is especially true when using mobile platforms where datalink and bandwidth issues are significant [12, 13].

In this study, the data to be acquired and processed by an intelligent measurement system comes from various sensors on-board a UAS, such as radar, electro-optical/infrared, GPS and Inertial Navigation Systems (INS). Apart from on-board input data streams, additional contextual input data can also be taken into account. The choice of which contextual input to apply can be automatically tailored using the computational intelligence techniques.

Four learning techniques are currently available within the IMS and are used for building the models – three of which are CI-based: artificial neural network (ANN), support vector machine (SVM), Bayesian network (BN), and K-means classifier. These techniques are implemented in JAVA and the Encog machine learning library [12]. Built in statistical analyses include Difference, Average, Variance, Standard deviation, Summation, Min and Max.

The simulated data set includes: X and Y locations of each vehicle on the road over the surveillance distance of a 6 kilometre road with three lanes, as well as the ground truth labels of driver types.

An inference model can be represented as a tuple $(\mathcal{L}, \overline{\mathrm{S}}, \overline{\mathrm{p}})$, where $\overline{S}$ is the vector of secondary measurands, $\overline{p}$ is a parameter vector of the learning technique $\mathcal{L}$, specifying such values as, for example, an error acceptance rate for artificial neural networks. The process of building an inference model is, in fact, an application of the learning technique $\mathcal{L}$ with its set of parameters $\overline{p}$ to the vector of chosen secondary variables $\overline{S}$ that provides both training and testing data inputs.

Having built the inference models corresponding to all the learning techniques used, this case study explores the influence of salient features of the modelled system on the performance of the IMS. As an example, one salient feature of the traffic simulation model is the ratio of abnormal and normal drivers, which in our experiments varies from 5 % to 25 %. The dependence of inference accuracy on this ratio for each learning technique implemented by the IMS are shown in Figs. 3 and 4.

Therefore, a multi-tiered IMS that uses computational intelligence techniques should be able to enhance situation awareness of a UAV, especially in a real-time environment. Once anomalies are identified from direct measurements, additional data from both easily accessible and detail-rich data sets can be added to improve the system classification and prediction performance.

For balanced training (Fig. 3), the numbers of training samples representing normal drivers is limited by the number of samples representing abnormal drivers of a particular type, which are relatively small in the case-study.

For unbalanced training (Fig. 4), the size of the training dataset representing normal drivers can exceed that of the abnormal ones. All other experimental parameters,

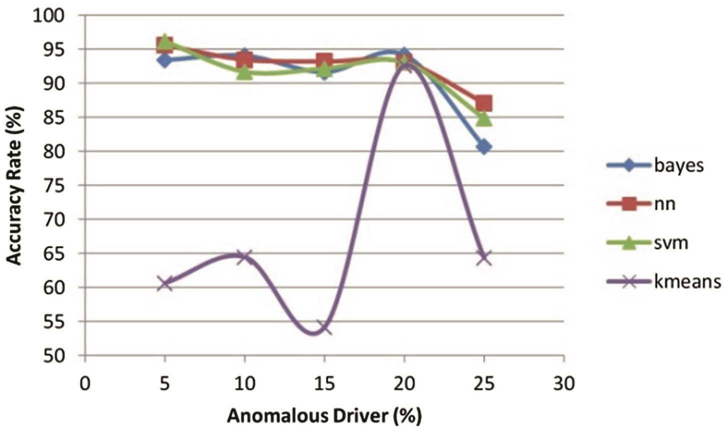


**Fig. 3.** Multiple data sources fused by an IMS

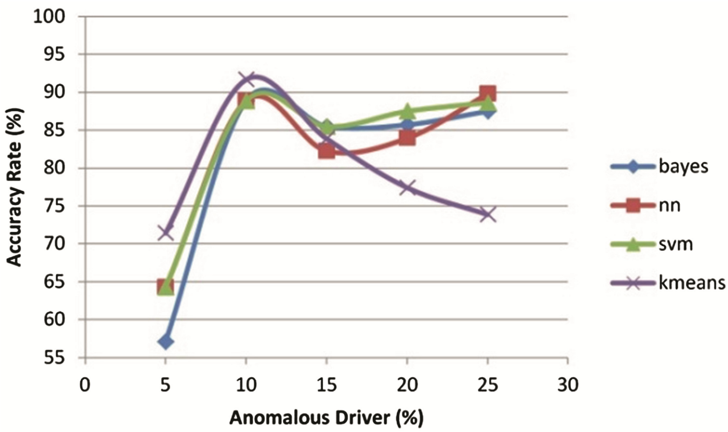


**Fig. 4.** The effect of different surveillance distances on system accuracy

including the size of testing datasets and the ratio of normal vs. abnormal drivers, are the same.)

As can be observed from the two figures above, the performance of supervised learning techniques is by and large similar, especially for smaller ratios of the numbers of abnormal and normal drivers, denoted as $\lambda_{ab/n}$.

SVM outperforms other supervised learning techniques when $\lambda_{ab/n}$ is small (<10 %), whereas for large values of $\lambda_{ab/n}$ (>20 %), ANNs become the best choice of supervised learning used for building inference models.

Unsupervised learning generally shows worse performance, but can also reach quite high inference accuracy. Despite their inconsistency in inference accuracy, unsupervised algorithms (unlike their supervised counterparts) do not require training. The ground truth labels obtained from MATLAB simulation (i.e. driver types – “normal” and "cowboy") are used only for validating these algorithms. This implies that the unsupervised algorithms converge much quicker and can be useful in cases when no (or very limited) training can be provided. It may also be possible to use an unsupervised algorithm as a precursory approach, while a training process of the supervised algorithms is carried out.

## 5 Conclusions

On the basis of the research work conducted in the present study, which was aimed at the development of IMSs for enhancing situation awareness of an UAS, the following conclusions can be drawn:

First of all, it has been shown that the concept of an IMS is viable in the chosen context – it has been demonstrated that the implementation of a framework for building such measurement systems is a feasible task, even with limited amounts of data available for making inferences.

Secondly, one of the main benefits of an intelligent measurement system, i.e. the ability to discover relationships between the primary characteristics of the system being monitored and the observed or measured data, has been demonstrated by inferring the behavioral type of drivers.

Thirdly, an essential step in building a good inference model is the selection of the most appropriate set of secondary measurands done semi-automatically by the proposed IMS that is achieved by adaptive filtering of input data streams.

Finally, the inference models within an IMS can be efficiently built with the help of machine learning techniques, which use both supervised and unsupervised approaches to learning. The ANN-based model of the process under observation proved to be the most adequate.

The experiments conducted on several simulated datasets and have demonstrated that reasonable performance can be achieved in terms of accuracy of data processing and its speed. For comprehensive evaluation of the developed IMS aimed at enhancing situational awareness of a UAS, however, it would be desirable to deploy the system on a mobile computing platform and to feed it with real-time sensor data, related to traffic surveillance. Experimenting with such a setup will inevitably bring some programming and engineering issues to the forefront, addressing which would reinforce system usability.

**Acknowledgment.** The authors would like to acknowledge the contribution of their industrial partner – Selex ES, a subsidiary of Finmeccanica Company – for providing the traffic simulation model, funding and general support for the work on this research project.

## References

1. Warne, K., Prasad, G., Rezvani, S., Maguire, L.: Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. Elsevier: Eng. Appl. Artif. Intell. **17**, 871–885 (2004). doi:10.1016/j.engappai.2004.08.020
2. Buyan, M.: Intelligent Instrumentation: Principles and Applications. CRC Press, Boca Raton (2010). ISBN 9781420089530
3. Khatibisepehr, S., Huang, B., Khare, S.: Design of inferential sensors in the process industry: a review of Bayesian methods. J. Process Control **23**, 1575–1596 (2013)
4. Sjoberg, J., et al.: Nonlinear Black-Box Modelling in System Identification: A Unified Overview. Automatica. Elsevier, Amsterdam (1995)
5. Khan, Z., Ali, A.B.M.S., Riaz, Z. (eds.): Computational Intelligence for Decision Support in Cyber-Physical Systems. Springer, Heidelberg (2014)
6. Rattadilok, P., Petrovski, A., Petrovski, S.: Anomaly monitoring framework based on intelligent data analysis. In: Yin, H., Tang, K., Gao, Y., Klawonn, F., Lee, M., Weise, T., Li, B., Yao, X. (eds.) IDEAL 2013. LNCS, vol. 8206, pp. 134–141. Springer, Heidelberg (2013)
7. Petrovski, S., Bouchet, F., Petrovski, A.: Data-driven modelling of electromagnetic interferences in motor vehicles using intelligent system approaches. In: Proceedings of IEEE Symposium on Innovations in Intelligent Systems and Applications, INISTA 2013, Albena, Bulgaria, pp. 1–7 (2013)
8. Lun, Y., Cheng, L.: The research on the model of the context-aware for reliable sensing and explanation in cyber-physical system. Procedia Eng. **15**, 1753–1757 (2011)
9. Park, K.J., Zheng, R., Liu, X.: Cyber-physical systems: Milestones and research challenges. Comput. Commun. **36**(1), 1–7 (2012)
10. Kromanis, R., Kripakaran, P.: Support vector regression for anomaly detection from measurement histories. Adv. Eng. Inform. **27**(2013), 486–495 (2013)
11. Valavanis, K.P.: Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy. Springer, Heidelberg (2007)
12. Encog Library. http://www.heatonresearch.com/encog