Predictive analytics methodology for smart qualification testing of electronic components

Stoyan Stoyanov, Mominul Ahsan, Chris Bailey, Tracy Wotherspoon & Craig Hunt

Journal of
Intelligent
Manufacturing

Volume 24   Number 4   August 2013   ISSN: 0956-5515

ONLINE FIRST

Available online
www.springerlink.com

🐴 Springer

🐴 Springer

Springer

CrossMark

Predictive analytics methodology for smart qualification testing of electronic components

Stoyan Stoyanov[1] · Mominul Ahsan[1] · Chris Bailey[1] · Tracy Wotherspoon[2] · Craig Hunt[2]

Abstract

In electronics manufacturing, the required quality of electronic modules (e.g. packaged electronic devices) are evaluated through qualification testing using standards and user-defined requirements. The challenge for the electronics industry is that product qualification testing is time-consuming and costly. This paper focuses on the development and demonstration of a novel approach for smarter qualification using test data from the production line along with integrated computational techniques for data mining/analytics and data-driven forecasting (i.e. prognostics) modelling. The most common type of testing in the electronics industry—sequentially run electrical multi-parameter tests on the Device-under-Test (DUT), is considered. The proposed data mining (DM) framework can identify the tests that have strong correlation to pending failure of the device in the qualification (tests sensitive to pending failure) as well as to evaluate the similarity in test measurements, thus generating knowledge on potentially redundant tests. Mining the data in this context and with the proposed approach represents a major new contribution because it uncovers embedded knowledge and information in the production test data that can enable intelligent optimisation of the tests' sequence and reduce the number of tests. The intelligent manufacturing concept behind the development of data-driven prognostics models using machine learning techniques is to use data only from a small number of tests from the full qualification specification as training data in the process of model construction. This model can then forecast the overall qualification outcome for a DUT—Pass or Fail—without performing all other remaining tests. The novelty in the context of machine learning is in the selection of the data features for the training dataset using results from tests sensitive to pending failure. Support Vector Machine (SVM) binary classifiers SVM models built with data from tests sensitive to the outcome that the module will fail are shown to have superior performance compared with models trained with other datasets of tests. Case studies based on the use of real industrial production test data for an electronic module are included in the paper to demonstrate and validate the computational approach. This work is both novel and original because at present, to the best knowledge of the authors, such predictive analytics methodology applied to qualification testing and providing benefits of test time and hence cost reduction are non-existent in the electronics industry. The integrated data analytics-prognostics approach, deployable for both off-line and in-line optimisation of production test procedures, has the potential to transform current practices by exploiting in a smarter way information and knowledge available with large datasets of qualification test data.

Keywords Data mining · Machine learning · Smart qualification testing · Electronic components · Intelligent manufacturing · Prognostics modelling

✉ Stoyan Stoyanov
  s.stoyanov@gre.ac.uk

1 Computational Mechanics and Reliability Group, School of Computing and Mathematical Sciences, Faculty of Liberal Arts and Sciences, University of Greenwich, Park Row, Greenwich, London SE10 9LS, UK

2 Microchip Technology Inc., Caldicot, Monmouthshire NP26 5YW, UK

Introduction

The global market for electronic products is projected to reach US$2.4 trillion per year by 2020 (Pecht et al. 2016). This growth has led to intense competition between manufacturers to minimise the time-to-market and cost of their products while at the same time delivering high quality and reliable products to their customer. Assuring the robust functional performance and quality of manufactured

∕ Springer

electronics products, and respective "fit-for-purpose" characteristics, requires the adoption of qualification processes, along with reliability testing, that often are time-consuming and resource-intensive (Ruidong and Chun 2017). A major challenge from economics point of view is the ability to reduce time-to-market for an electronic module that satisfies customer requirements in terms of quality and reliability. Identifying solutions of how to overcome this challenge is a high priority for electronics manufacturers.

Qualification, being part of the production line of a product, is an application-specific process. In general, manufacturers develop qualification specifications for electronics products based on respective application requirements, as defined by the customer, and industrial standards. Tests (Qualification and Environmental) are used to determine whether the product meets the specified requirements in terms of quality and reliability before it is delivered to the customer (Wang et al. 2008; Vichare et al. 2006).

Qualification tests of electronics products are conducted typically through measurements of various electrical parameters that are indicators of the functional state of the individual electronic component or product. A qualification test outcome is typically binary and defined as either PASS or FAIL based on the measured test values and associated specified test limits. A test value within the expected test range is associated with pass status and indicates that the required quality is present. It is a common practice in the electronics industry to archive qualification test measurements, for example to ensure traceability information is available, and as a result manufacturing enterprises often have access to large historical sets of test data for their products. This data often stays unused but can potentially hold valuable information and knowledge that can enable the optimisation of the test procedures required for the respective product manufacturing line.

Data mining (DM) and machine learning (ML) offer a powerful approach to problems in manufacturing that require extracting information and decision making by means of data analytics and predictive modelling. However, current use of computational intelligence and data for enabling smart manufacturing solutions remains limited in many industrial applications. In recent years, as result of the increasing use, adoption and advances in Internet of Things (IoT) and Information and Communications Technology (ICT), this has started to change. IoT and ICT provide huge opportunities for the paradigm of Industry 4.0 and smart manufacturing. Production data from diverse sources and in various formats are gathered and stored in increased volumes, and is rapidly turning into one of the main pillars of smart manufacturing (Kusiak 2018). This data can be used to develop and imbed intelligence and smartness in manufacturing. Being in possession of large databases gathered from their production floors and machinery, more and more manufacturing

enterprises are starting to recognise the importance of adopting data mining (DM) approaches in order to increase their own competitive advantage by exploiting the information and knowledge imbedded within manufacturing datasets (Wang 2007). However, generating knowledge from large datasets and the analytical tools required to do that remains a key challenge.

Data mining for monitoring and improving the quality of high-tech manufacturing has gained attention in the past years and has been researched increasingly with respect to a range of challenges that datasets and applications pose. New intelligent data analytics methodologies that can enable the automation of the information and knowledge extraction are required to deal with large complex datasets with records and features gathered on modern production lines (Choudhary et al. 2009). Looking at the nature and implications for applying data mining in manufacturing, Wang points at several factors that have major influence on the success for the adoption of data mining (Wang 2007). Among these, availability of appropriate datasets, data cleaning and pre-processing, selection of suitable features for the knowledge to be solved, and machine learning model capability to identify correlation relationship in the data are specified as being very important.

Published research provides some interesting examples of data mining and machine learning applications for intelligent manufacturing covering problems related to quality control, scheduling, fault diagnosis, defect analysis, decision support systems, etc. as well as the development of DM approaches and algorithms (Braha 2013). Research by Chien et al. (2014) focused on the integration of design of experiments with data mining for knowledge extraction that supports accurate characterisation of the yield performance of newly released wafer fabrication technologies as well as diagnosis in the instance of large datasets automatically collected in semiconductor manufacturing. Wuest et al. (2014) proposed a quality control approach based on machine learning techniques and clustering analysis for modern manufacturing programmes that generate product state data with increasing complexity and high-dimensionality of the features. Rokach and Maimon (2006) have developed a feature set decomposition methodology for applications concerning quality improvement through classifications models. An interesting research question tackled in their work is the handling of data in cases when the training set size of the data is small relative to the number of the features. There has also been research on the adoption of manufacturing equipment data for building predictive models for condition monitoring and prognostics. For example, Benkedjouh et al. (2015) utilised methods for nonlinear feature reduction and support vector regression using data on monitored signals from high precision cutting tools to assess their wear evolution/degradation and to predict the tool remaining useful life (RUL).

Da Cunha et al. (2006) detail an interesting application of data-mining approach using production data for the reduction of the risk of producing faulty products by identifying an optimal sequencing of assembly tasks. More recently, Godreau et al. (2018) demonstrated a new data mining method using unsupervised learning for the classification of equipment critical events and their aggregation. The authors have also researched the potentials of contextual clustering for a better data selection.

Applications of data mining and machine learning in electronics manufacturing domain become more common and increasingly important as companies recognise that such smart expert systems and imbedded data-driven intelligence can provide competitive advantages in a global economy. Published research has addressed manufacturing challenges related to fabrication, yield optimisation, and automated fault/defect detection (Stoyanov et al. 2016; Park et al. 2013; Sohn and Lee 2012; Kupp and Makris 2012; Kim et al. 2012, 2015; Chou et al. 1997; Boubezoul et al. 2007).

The research reported in this paper aims at the formulation, demonstration, and validation of a data mining/machine learning methodology for smart qualification testing of electronic products. The type of qualification specs targeted with this work is the common electrical parameter testing where hundreds of individual tests are executed sequentially, one after the other. The proposed data analytics-based modelling methodology can provide insights into the role and significance of individual tests and their sensitivity to being reliable precursors of pending failure of the qualification process. The novelty and contribution is in the data mining approach underpinned by data distribution modelling and the proposed similarity assessment method for qualification tests, identifying sensitive to pending failure tests from the full set of qualification tests, and then use this data to develop more efficient and accurate predictive models through the selection of a smarter training dataset. Uncovering this knowledge from the data results in the following two opportunities/objectives: (1) Optimise the production test specification for electronics DUTs, and (2) Develop an intelligent manufacturing capability for qualification testing through in-line imbedded, model-based, prognostics. The qualification process can be optimised by identifying favourable sequencing of the individual tests and if there are any potentially redundant tests that are not required. Also, test time reduction can be achieved by the proposed in-line adoption of data-driven, machine learning prognostics models that can provide, with a degree of accuracy, predictions for the expected overall qualification outcome (Pass or Fail) for a DUT without executing all tests in the qualification spec.

The computational approach and the associated methods used to mine the data and to develop the ML prognostics model in relation to the two objectives above are demonstrated and validated using real production test data gathered on the qualification of an electronic module.

Opportunities with historical qualification data

The type of qualification testing considered in this work requires undertaking a sequential series of electrical parameter measurements on the electronic device. In a sequence of individual tests, for the device to be qualified, it is required that all individual tests constituting the qualification specification are passed. The measured test value has to be within a predefined range in order for that test to be passed; otherwise, the DUT fails the specific test, and respectively the whole qualification. When a DUT fails a test in the test sequence, continued testing of that device is stopped.

Many qualification procedures of this type require a large number of electrical, logical and other functional parameter measurements, which for complex electronic parts can easily require individual tests in the order of hundreds. Hence, the overall qualification of a single electronic part can easily become time consuming given the need to perform electrical probing for such a large number of parameters.

Given the sequential approach for executing the individual tests, a potential way to shorten the overall qualification time is by taking advantage of the fact that as the testing progresses, more and more data of completed individual tests becomes available. Mining and analysing this data, and forecasting the overall qualification outcome with prognostics models is an appealing prospect that can enable reducing the required number of tests (e.g. through redundant tests identification and/or optimal tests sequencing) and hence time-to-market and cost. In essence, there is a clear opportunity to build machine learning models using past historical data on qualified electronic devices, and then embed the models in the qualification process for in-line use to forecast the qualification outcome.

The remaining sections of the paper detail the data-driven modelling methodology for smart qualification testing that is proposed, developed and demonstrated using real qualification test datasets for an electronic module.

Methodology and computational approach

The proposed methodology is developed to take advantage of the availability of large historical qualification test datasets that many manufacturers in the electronics industry have generated and archived over long periods of time but have not exploited yet to achieve more intelligent production lines. The developed computational approach is capable of supporting smarter testing through optimisation of qualification

Fig. 1 Approach to smart qualification test of electronic products
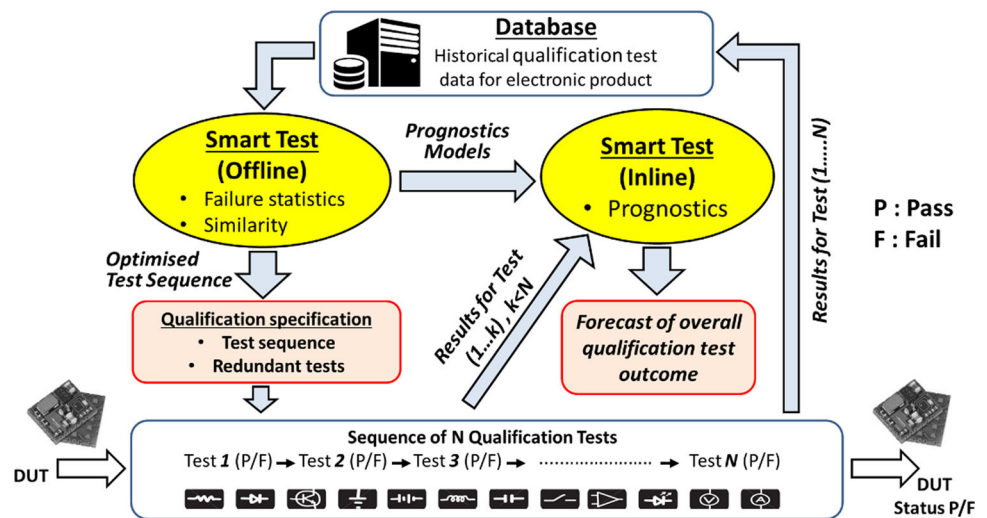
test specifications using the knowledge obtained through data mining and data analytics of the test data, and by constructing and applying data driven prognostics models to forecast expected qualification outcome for the tested device. The big picture of the smart-test strategy is detailed in Fig. 1. Results from failure statistics and similarity test evaluation can be derived offline by accessing historical datasets. This information can then be used for qualification optimisation: suggesting a different order of test execution and identifying what might be the potentially redundant tests. The other key opportunity with offline mining and analytics of the test data is the generation of prognostic models which can be used to forecast the output of subsequent tests. The approach is to base this on test measurements from a small number of individual tests completed on a DUT, and predict the overall final qualification status without executing the remaining tests beyond the point of the current test at which the model-based forecast is made.

The developed numerical modelling approach is based on integrated techniques for statistical analysis of failure test data, distribution-based data modelling, and data mining/analytics for identification of (1) potentially redundant tests and (2) tests sensitive to pending failure. Machine learning based modelling for in-line qualification outcome prognostics that takes advantage from the identified tests sensitive to pending failure is also incorporated within the computational framework.

Figure 2 illustrates our concept of bringing information from data mining, particularly on identified so-called "sensitive to pending failure" tests, to support the adoption of more efficient and accurate machine learning prognostic models within the qualification process execution. As detailed in Fig. 2, we first perform a set of data pre-processing tasks on the raw test data recorded: cleaning, formatting, and normalisation of the data. This is undertaken before the data

is subjected to data mining and analytics. Combined distribution and similarity modelling, along with failure statistics, are used to generate knowledge on how the qualification process can be optimised. Identification of potentially redundant tests from similarities in PASS data distributions with zero failure statistic can enable test time reduction of the overall qualification process for DUTs by removing those tests for the qualification specification.

Performing similarity analysis, but this time on data for a given test gathered separately on PASS and FAIL devices, identifies tests that are sensitive to pending failure. This is regarded as the most novel aspect of the proposed computational approach in the context of designing the training datasets required for the machine learning classification model developments. Training data that includes tests sensitive to pending failure will improve the accuracy of the constructed ML models while simplifying their complexity (i.e. less model input information). Predictive models for the final PASS or FAIL qualification outcome can then be developed from the test data available at any given test in the sequence, and used at that point to forecast if the DUT will be successfully qualified. Decision to stop or continue the sequence of tests can then be taken in the context of the known model accuracy as well as the application requirement for yield.

To facilitate the understating of the proposed methodology and algorithmic framework, a detailed block diagram of the computational steps required with our approach is produced. This diagram is detailed in Fig. 3. The building blocks for this approach require performing the following main steps:

Step 1   *Obtain dataset* Access historical qualification test database for a given electronic device and application. The type of qualification is the most common for the electronics industry multiple parameter elec-
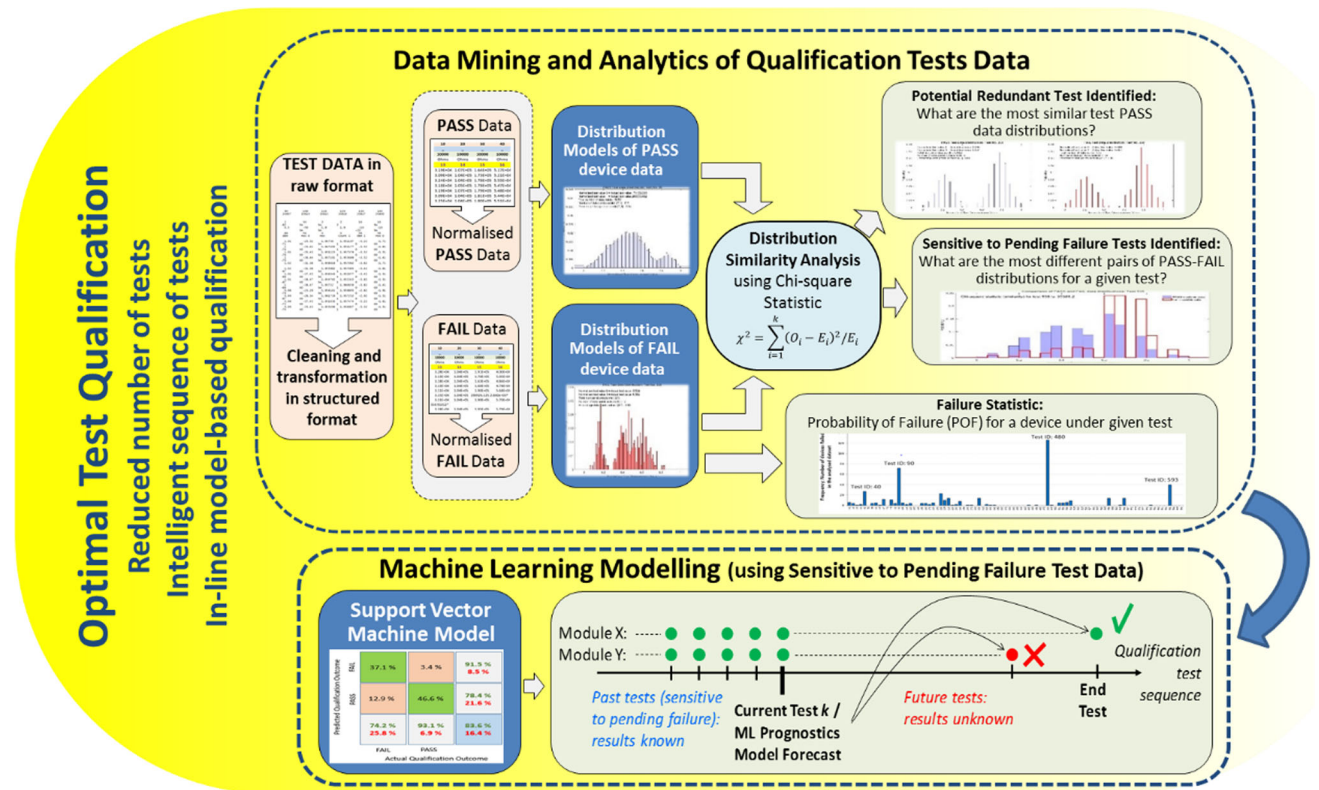
Fig. 2 Integration of the developed data analytics approach for optimisation of qualification testing of electronic devices with Machine Learning (ML) prognostics models for forsaking qualification outcomes for DUTs

trical/functional test measurements. We adopt the abbreviation DUT (Device-under-Test) for the electronic device of interest.

Step 2 *Cleaning and pre-processing* Cleaning and pre-processing of the raw. At this step, the raw data (test device records N_r, and number of individual tests in the sequence is d_r) needs to be transformed into clean data that is suitable for data mining and analytics. Complete data fields such as those that may hold textual test related information or symbols acting as labels for certain test attributes, for example a * symbol marking a test measurement outside the test limits, etc., are removed in the produced clean dataset. Cleaned data must be numeric only; thus, only tests with outcomes that can be transformed into numerical results are processed. PASS and FAIL status are coded as − 1 and 1 respectively. Tabular structure for the cleaned data is adopted: (1) each row contains the numerical test data for one device and (2) each column corresponds to a test that is performed. Hence, a field (cell) in this tabular structure, for example implemented in CSV or Excel file format, holds the measurement for a given device under given test. The columns, from left to right, follow the same order as the sequence of the test in the actual

procedure. If a device has passed the qualification, all individual tests have bene passed and measurements for all test values are included (a complete set of data in all cell in tabular row for that device). If a DUT has failed, then test measurements in the row for that device appear in the cells up and including the column for the test under which the failure has occurred, with all remaining cells in that row on the right of the failure test column being empty. This reflects on how such qualification tests are most commonly undertaken by the industry. Once a device fails a test in the sequence, testing of the device is terminated and none of the remaining tests is performed.

In practice, a device that has failed the qualification may be re-tested one or more times (for example if the DUT has not be seated properly on the test bench). If multiple records for the same device are stored as a result, as part of the cleaning the last stamp test record is maintained, with all other test attempts removed from the dataset.

It is the user's responsibility to handle the raw data cleaning and achieve data transformation into a tabular structure as outlined in this Step (2) and using a suitable data file format to store the clean data. Remaining steps are performed on the cleaned
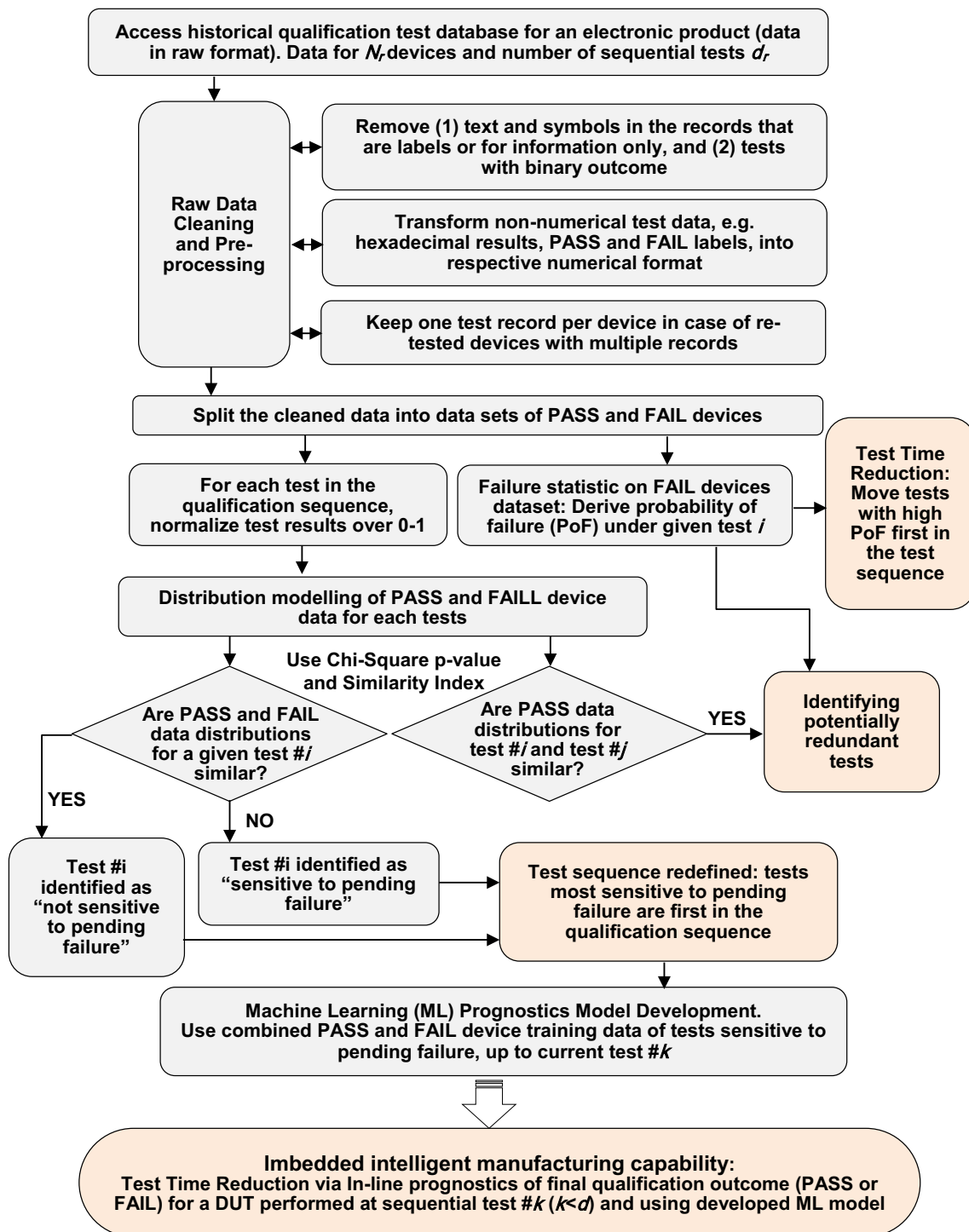
```
┌─────────────────────────────────────────────────────────┐
│ Access historical qualification test database for an     │
│ electronic product (data in raw format). Data for Nr     │
│ devices and number of sequential tests dr                │
└─────────────────────────────────────────────────────────┘
```

Access historical qualification test database for an electronic product (data in raw format). Data for N_r devices and number of sequential tests d_r

Raw Data Cleaning and Pre-processing

Remove (1) text and symbols in the records that are labels or for information only, and (2) tests with binary outcome

Transform non-numerical test data, e.g. hexadecimal results, PASS and FAIL labels, into respective numerical format

Keep one test record per device in case of re-tested devices with multiple records

Split the cleaned data into data sets of PASS and FAIL devices

For each test in the qualification sequence, normalize test results over 0-1

Failure statistic on FAIL devices dataset: Derive probability of failure (PoF) under given test i

Test Time Reduction: Move tests with high PoF first in the test sequence

Distribution modelling of PASS and FAILL device data for each tests

Use Chi-Square p-value and Similarity Index

Are PASS and FAIL data distributions for a given test #i similar?

Are PASS data distributions for test #i and test #j similar?

YES → **Identifying potentially redundant tests**

YES

NO

Test #i identified as "not sensitive to pending failure"

Test #i identified as "sensitive to pending failure"

Test sequence redefined: tests most sensitive to pending failure are first in the qualification sequence

Machine Learning (ML) Prognostics Model Development. Use combined PASS and FAIL device training data of tests sensitive to pending failure, up to current test #k

Imbedded intelligent manufacturing capability: Test Time Reduction via In-line prognostics of final qualification outcome (PASS or FAIL) for a DUT performed at sequential test #k ($k<d$) and using developed ML model

Fig. 3 Block diagram of the proposed data analytics approach for smart qualification testing of electronic devices

dataset where the number of tested devices is N ($N<N_r$) and the number of individual tests maintained is d ($d<d_r$).

Step 3 *Predict probability of failure* The actual clean test data, containing records for both PASS and FAIL DUTs, is first separated into PASS and FAIL

datasets. Failure statistics for all tests is undertaken using the available data, and values for probability of failure (POF) for a device under given test are obtained.

Step 4 *Normalise the data* The test data is normalised over the range 0–1. This is an important step to enable

subsequent data analytics, specifically in the context of similarity comparisons, and data-driven model development. Different numerical strategies for the normalisation of the data can be adopted, some of these are discussed in greater detail in Sect. 4.2.1. For example, a possible scheme to normalise the data is the use the PASS data for a given test and use the 5th and 95th percentile of the actual data as 0 and 1 values respectively in the data normalised space for that test.

Step 5 *Construct probability distributions* The normalised test data of PASS and FAIL DUT datasets are used to derive, for each of the individual tests in the spec, distributions of the respective test measurements in the format of probability distributions (histograms). It is important that for all tests, constructed histograms are specified to have exactly the same number of bins in the normalised interval 0–1. This a mandatory requirement to enable the similarity calculations detailed with Steps 6 and 7.

Step 6 *Chi square test statistic* The Chi square test statistic and goodness-of-fit p value are used to calculate, using the PASS data distribution for each test, how similar the data distribution of that test is to the distributions of test data gathered from all other qualification tests. The Chi square test statistic calculation using histograms is detailed in Sect. 4.3.1. This theory is used to propose and adopt the use of a metric, so-called Similarity Index (*SI*), that can rank data distributions, and in this instance respectively qualification tests, based on their similarity. The definition of the *SI*, along with the conventional p-value from the Chi square goodness of fit test, are summarised in Sect. 4.3.2. Mining of test data for similarity in test data distributions enables to identify groups of tests that can be seen as playing similar role in the qualification of the device. This knowledge, along with the test failure statistic information, can be used to support identifying (potential) redundant tests. While engineering judgment is needed, and application-specific requirements have to be accounted for, from data analytics point of view an individual test can be considered as a candidate for redundancy if the following two test attributes exists simultaneously:

- No, or near zero (if acceptable), failures of the device under the test, and
- The test outcome, in terms of distribution of the result values, follows similar distribution as the distribution of one or more preceding tests. The criteria for minimum level of similarity is based on having p value greater than 0.99, where the p value is calculated from the Chi square goodness

of fit test applied to the PASS data distributions of the two tests.
.

Step 7 *Similarity index* The Chi square statistic is used, similarly as in Step (6) above, to evaluate the similarity (respectively dissimilarity) in the measured test data in the instance of PASS and FAIL DUT data for a given test. This analysis informs which tests are potentially sensitive (or not sensitive) to pending failure. If the distributions of measured test data from PASS and FAIL devices, for a given test, differ notably then the test has the potential to detect out-of-the-norm device performance that may be associated with pending failure under a subsequent test in the remaining part of the qualification test sequence. The results can be used to inform on the existence and the potential of qualification tests to underpin the construction of predictive machine learning and fault classification models for in-line test prognostics. The ranking of the tests in the qualification in terms of their sensitivity to pending failure uses the computed values for the proposed Similarity Index defined in Sect. 4.3.2.

Step 8 *Optimise sequence of tests* Optimisation of qualification test sequence requires tests with high failure rate and/or tests sensitive to pending failure to be performed first in the sequential testing process. Likely failure under tests in the sequence undertaken first in the sequence of tests can offer reduction in test time by avoiding unnecessary near-zero failure rate tests (because they come later in the sequence). Also, machine learning techniques can be used to develop prognostic models that need only limited number of completed tests, those undertaken first and sensitive to pending failure, and offer predictive accuracy that is superior compared with other test data use strategies.

Step 9 *Build training dataset and prognostics model* Decide on a test in the optimised test sequence (denote test # k) at which a ML model will be applied to forecast the likely outcome of the overall qualification—pass or fail. Use historical test data on the first k tests in the optimised qualification spec to create a train in dataset. Train a model structure of a binary classifier (for example Support Vector Machine) to obtain a prognostics model. Imbed the prognostics model for in-line evaluation of qualification of a DUT.

The programming code for the implementation of the methods and calculations associated with the computational steps 1–9 of the proposed data analytics approach, as detailed above, is written in MATLAB scientific computing pro-

gramming environment (version R2018). The development (writing, debugging and commenting) of the programming code for the approach detailed in this section, and in the context of this published work, is referred to as *alpha-stage* software development and has taken approximately 150 person-hours of effort.

The current *alpha-stage* version of our MATALB programming code that realises all required calculations, and to obtain the results reported in this paper, has 1550 lines of scripting. This number includes also the lines of comments on the code that have been added. Follow up work that authors are planning will aim to target specifically the transition of the programming code to *beta-stage* version that will be optimized version of the current code, comprehensively documented, and considered, subject to agreement and approval by the funder, for release as open-source upon completion.

Qualification test data modelling

Qualification datasets used in the study

The main feature of the analysis approach developed/applied to the qualification test data is that it is a numerical approach. For this reason, individual tests in the qualification procedure are treated as being equally important and equally significant in qualifying the electronic module and determining the final overall qualification outcome (PASS or FAIL). The data is in the format of structured numerical data gathered through performing a series of sequentially executed test measurements. Results obtained with this approach have to be considered in addition to applying appropriate engineering judgement (e.g. using knowledge about the role of a test in the qualification, the physical aspect of product testing, etc.) and used in the in the context of the respective product and application.

Historical qualification test datasets for 50,000+ electronic modules, referred here as Device under Test (DUTs), are investigated. The proposed numerical approach for data analytics is applied to assess a qualification procedure that encompasses 150+ individual sequential tests. Some of the tests are measurements that have the test outcome as real value numbers, for example voltage, current, time durations, power ratios, signal power strength and frequency. Other tests provide measurements in Hex units, which, once transferred from hexadecimal to decimal numbers, result in an integer number. There are also tests for which the test parameter is integer, for example, those that provide a count of some test related characteristics. There are also logical tests that output *True* or *False* values. Hence the sequence of tests result in datasets that are real, integer and logical.

Some of the tests have double sided limits for the PASS test condition and others are single sided. There are also test

results only provide information and hence do not affect the qualification status and therefore are ignored in this analysis.

Performing a numerical-based analysis for such a range of diverse tests that follows a generic (non-specific to the test) computational approach is challenging. Following preliminary investigations of the datasets, it was decided to develop the data analytics approach based on distribution modelling of the qualification test data and mining the data behaviour/relationships using suitable techniques. Such an approach can offer robustness and generalisation of the proposed computations that are judged as being the most important attributes of the proposed Smart-Test framework.

Considering test data behaviour via the distribution of data is meaningful only for the qualification tests that generate varying results. Therefore, the data mining studies are undertaken only on a subset of tests for which the test result varies and can be modelled as a distribution. For the data investigated, there are 111 qualification tests, out of the total 150+ tests, that meet this criteria. All following studies detailed in the paper use test data that is gathered only from these 111 tests. Remaining tests which are of the type logical tests with PASS condition requiring TRUE (or FALSE) test outcome and match hexadecimal (HEX) tests with PASS condition requiring a specific Hex value, are excluded from the datasets handled with the proposed approach, and thus have no influence on obtained results.

Figure 4 shows a simplified, illustrative sample of the raw measured tests data, in the format of normalised values, for a particular type of electronics module. The measured parametric values for each sequential test are arranged column wise, and the test results for each electronic module appear in a row of the presented table. It should be noted that no further tests are carried out once a module has failed under a particular test in the test sequence. Hence, no test data will be available for a module onwards from the test of failure. The first four rows show indicative information about parametric tests number/parameter, and upper and/or lower limits for the PASS criterion of each individual test.

Distribution modelling of test data

Cleaning and normalization of test data

Measurements from different qualification tests are different. Some measurements are numerical decimal continuous values, others give the result as an integer number or as a Hex value. There are logical tests too. The magnitude/order of the measured value (where numerical) can also be very different. Measurement units from test to test change. Some tests are double side limited, some have a limit only on one side. The best strategy in numerical analysis to handle such differences is to subject the data to normalisation. The nor-

						Test Number				Known outcome
	1	2	3	4	....	10	11	....	N	
Upper limit	1	1	1	1	....	1	1	....	1	
Lower limit	0	0	0	0	....	0	0	....	0	
Module 1	0.11897	0.88349	0.48914	0.53865	....	0.18535	0.48090	....	0.02012	PASS
Module 2	0.08448	0.99432	0.55154	0.56794	....	0.17126	0.57001	....	0.02273	PASS
Module 3	0.08046	0.91191	0.51449	0.53167	....	0.14965	0.65764	....	0.01976	PASS
Module 4	0.03448	0.90622	0.41700	0.43496	....	0.16493	0.54107	....	0.02335	PASS
Module 5	0.13594	0.88349	0.18204	0.22847	....	4.11831				*FAIL*
Module 6	0.08851	0.99432	0.55154	0.57819	....	0.18305	0.57365	....	0.02172	PASS
Module 7	0.08621	0.91191	0.49792	0.50350	....	0.18197	0.65706	....	0.01984	PASS
Module 8	0.08793	0.99147	0.52619	-0.92373						*FAIL*
Module 9	0.09195	0.99147	0.45112	0.51727	....	0.12586	0.57273	....	0.02080	PASS
Module 10	0.12586	0.88349	0.48524	0.54832	....	0.20171	0.47633	....	0.02104	PASS
Module 11	0.03736	0.90338	0.43552	0.45752	....	0.12853	0.53894	....	0.02517	PASS
Module 12	0.09195	0.99147	0.50179	0.54334	....	0.16112	0.57304	....	0.02041	PASS
Module 13	0.08793	0.91191	0.52814	0.54480	....	0.16745	0.65678	....	0.02014	PASS
Module 14	0.12701	0.88349	0.51449	0.56677	....	0.22218	0.47361	....	0.02074	PASS
Module 15	0.03966	0.90338	0.42187	0.42354	....	0.15133	0.54013	....	0.02167	PASS
Module 16	0.09195	0.99147	0.49283	0.53279	....	0.11682	0.57115	....	0.02096	PASS
Module 17	0.08851	0.91191	0.45405	0.46513	....	0.16317	0.65714	....	0.02278	PASS
Module 18	0.12644	0.88349	0.44430	0.52459	....	0.19251	0.47399	....	0.02379	PASS
Module 19	0.07759	0.88633	0.51254	-0.99374						*FAIL*
Module 20	0.09310	0.99147	0.49889	0.52723	....	0.17413	0.57326	....	0.02128	PASS
Module 21	0.08908	0.91191	0.47939	0.49403	....	0.15793	0.65560	....	0.02003	PASS
Module 22	0.12471	0.88349	0.37313	0.48915	....	0.20590	0.47338	....	0.02207	PASS
....	....	....	....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....	....	....	....

Fig. 4 Illustrative example of qualification test data in normalised format

malisation scheme used in this study transforms the data using normalised limits of 0 and 1.

Accounting for the overall approach and the need for robust data handling, the following normalisation strategy is formulated and implemented:

1. The raw test data is first cleaned (for details refer to Sect. 3, Step 2 of the Methodology) and filtered into two, PASS and FAIL, datasets.
2. The data for DUTs that PASS the qualification is associated with the "normal" expected test behaviour in the context of measured values. For each test, different strategies to decide on the actual lower and upper limit values for the data normalisation over 0–1 have been considered and tested. The method we applied uses values for normalisation selected as a percentile of the entire PASS data set for a given test; this method enables a robust solution to the data normalisation problem. The low limit of the data was selected as the 0.1 percentile and high limit as the 99.9 percentile of the data. This way measurements with outlier characteristics were not ignored. Both PASS and FAIL data has been normalised over 0–1 range using the above explained percentiles as the actual limits for data normalisation.

Data distribution modelling

Modelling the distributions of the normalised data for each test provides details on the behaviour of the data (how the data is spread, the nature and magnitude of variation, etc.) and allows comparison between respective test data. As we deal with finite in size datasets of numerical values, the histogram modelling approach of data distribution is utilised. Normalised test values less than 0 are binned in a single bin and similarly a single bin holds all values above 1. A detailed distribution is generated within the 0–1 interval. Histograms generated on all tests use the same number of bins for the data over the normalised range 0–1. This is an important condition that the proposed approach demands in order to enable the subsequent use of the method for similarity evaluations of tests. The vertical axis of the histogram denotes 'probability'. The height of each bar is the relative number of observations (number of observations in bin/total number of observations), and the sum of all bar heights in the diagram is 1.

Figure 5 shows an example of histogram models for the data gathered on a particular qualification test. The DUTs that generated this data are devices that have passed all tests in the sequence up to and including the particular test for which
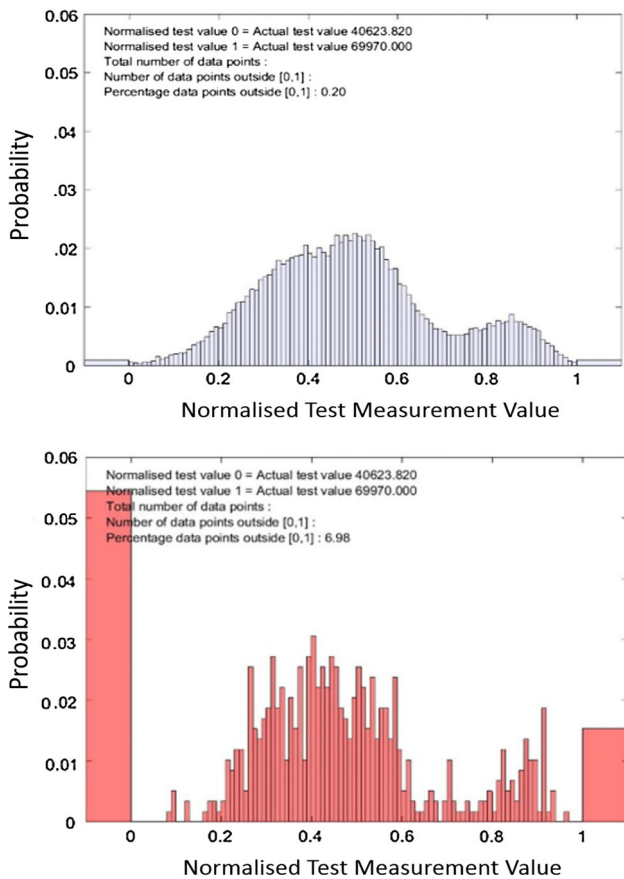
Fig. 5 Example of data distribution of measurements for a given test gathered from PASS-status electronic modules (top) and FAIL-status electronic modules (bottom)

this example is generated. The two presented histograms in Fig. 5 are built by splitting the test data into two subsets: (1) data from DUTs that passed all remaining tests in the qualification sequence and thus have overall PASS status, and (2) data from DUTs that have failed one of the remaining tests in the qualification sequence and thus have end-of-qualification FAIL status. The resulting distribution from the data in (1) above will be referred to as *Pass Test Data Distribution* (the histogram at the top of Fig. 5 test example) and the one from data in (2) as *Fail Test Data Distribution* (the histogram at the bottom of Fig. 5 test example). We explore this type of histogram distribution pairs, across all individual tests, in subsequent similarity evaluations and assessing the sensitivity of a test to pending failure.

In the case of the test data from an electronic module qualification demonstrated in this work, *Pass Test Data Distributions* for the tests in the qualification use data from over 50,000 tested modules. With such a large number of data points, the confidence in the obtained distributions is high. The size of the datasets underpinning the construction of the *Pass Test Data Distribution* is substantially smaller, in the order of 600.

Data similarity assessment

The use of histograms is convenient as it provides the ability to compare different qualification tests, and also PASS and FAIL data for a given test. By comparing how similar or different are two data distributions, important observations and conclusions regarding a qualification test procedure can be made. The quantitative approach to similarity assessments of data distributions uses the use of Chi square statistic.

In statistics, the Chi square goodness-of-fit test is often used to test if a sample of data comes from a population with a known distribution. An attractive feature of the Chi square test is that it can be applied to any univariate distribution for which the cumulative distribution function can be calculated. The Chi square goodness-of-fit test is always applied to binned data. In the case of non-binned data, one can simply calculate a histogram or frequency table before generating the Chi square test. As in our case the distributions are already in the format of histograms, the application of Chi square statistic technique is very straightforward.

Chi square test statistic

With standard use of Chi square test goodness-of-fit, the hypothesis that a set of data (i.e. observed values) comes from a population with a given (specified) distribution (i.e. expected values) is tested. This assessment uses the Chi square test statistic χ^2 that is defined as.

$$\chi^2 = \sum_{i=1}^{k}(O_i - E_i)^2/E_i \tag{1}$$

where O_i is the observed frequency of data for bin i and E_i is the expected frequency for bin i. The sum in Eq. (1) is over the non-empty bins (k). When using Chi square test for goodness-of-fit evaluation, expected frequency refers to the given distribution and is calculated as

$$E_i = N(F(Y_U) - F(Y_L)) \tag{2}$$

where F is the cumulative distribution function for the distribution being tested, Y_U is the upper limit for bin with index i, Y_L is the lower limit for data bin i, and N is the sample size of the observed data.

The value of the Chi square test statistic is dependent on how the data is binned and sensitive to the choice of the bins. There is no optimal choice for the bin width (since the optimal bin width depends on the distribution). Most reasonable choices produce similar, but not identical, results. A disadvantage of the Chi square test is that it requires a sufficient sample size in order for the Chi square approximation to be valid. In statistics, this will typically be posed as a requirement to have a minimum of 5 data values in each non-empty

bin. This has been the case with the data used in our demonstration case study.

The goodness-of-fit part of the computation is based on the fact that χ^2 follows approximately Chi square distribution with $(k\text{-}c)$ degrees of freedom where k is the number of nonempty bins and c is the number of estimated distribution parameters ($c = 1$ in this study). With a specified significance level α, a Chi square critical value ($\chi^2_{1-\alpha,k-c}$) is obtained using the Chi square distribution with $(k - c)$ degrees of freedom. The test statistic and the critical value are used to check the condition

$$\chi^2 > \chi^2_{1-\alpha,k-c} \tag{3}$$

If this relation is true then the hypothesis that the data are from a population with the specified distribution is rejected.

Similarity index (SI) and chi square *p* value

In this work, we adapt the use of Chi square test statistic calculation to meet our objectives. We require a similarity measure, or similarity index, that shows how similar is the test data distribution given with one histogram to the data modelled with another histogram.

First, the use of simple metric termed Similarity Index (*SI*) is proposed. The *SI* is the Chi square statistic value χ^2 normalised to the Chi square critical value $\chi^2_{1-\alpha,k-c}$ assuming significance level $\alpha = 0.01$ and $(k - 1)$ degrees-of-freedom where k is the number of non-empty bins in the histogram pair:

$$SI = \frac{\chi^2}{\chi^2_{1-\alpha,k-c}} \tag{4}$$

Larger values of S*I* indicate less similar datasets for a pair of histograms while small values of *SI* are indicators of greater similarity in the respective datasets. In this study, the proposed similarity index is used to rank the qualification tests given similarity between their respective PASS and FAIL data distributions. An example of similarity evaluation between data from PASS and FAIL devices for a test, with SI = 3.72, is illustrated in Fig. 6.

While the use of the defined Similarity Index offers a simple way to rank pairs of data according to the level of their similarity, this metric will not answer the question if a hypothesis that two datasets come from the same distribution can be accepted. This type of question comes into play when performing the study on identifying redundant tests. The standard Chi square goodness of fit test is using the so called p-value to accept or reject the H$_0$ hypothesis:

H$_0$ The sample data from test i follows the specified distribution for test j.
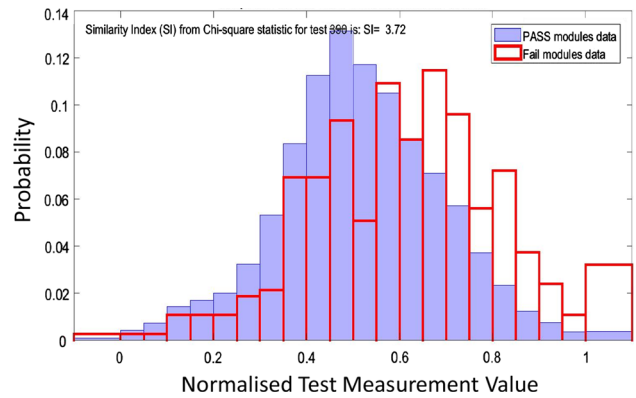

Fig. 6 Example of data similarity for a given qualification test in the spec, along with calculated Similarity Index (SI)

The *p* value is calculated as:

$$p \text{ value } = 1 - ChiSqCDF(\chi^2, DOF) \tag{5}$$

where *ChiSqCDF*() is the Cumulative Density Function (CDF) of the Chi Square distribution with *DOF* degrees of freedom. The *p* value is between 0 and 1. With significance level α (e.g. $\alpha = 0.01$), *p* value greater than α is the condition for the H$_0$ hypothesis to be accepted. The significance level is the probability that the hypothesis H$_0$ is rejected while it is actually true.

When we adopt the use of *p* value in this study, *p* value greater than 0.99 is used to make the assertion that two qualification tests are sufficiently similar in the context of redundant test identification, i.e. sample data from one test follows the distribution of the other test.

Sensitivity of tests to pending failure

Little or no similarity between a pair of *Pass Test Data Distribution* and *Fail Test Data Distribution* of a given qualification test is an indicator that the test produces measurement for the DUTs that can be used as a precursor for the final, overall qualification outcome, and thus can be used for prognostics. By calculating the Similarity Index for each pair of *Pass and Fail Test Data Distributions* using the respective data histograms, across all analysed qualification tests, it is possible to rank the tests with regards their data distribution similarity and this way to assess their sensitivity to detect pending failure.

Figure 7 shows two representative tests with dissimilar pairs of PASS and FAIL data distributions. The difference between the PASS data and the FAIL data distributions is illustrated by overlapping the two histograms and the value of the *SI* is included with each graph. Larger *SI* value means greater difference in the test PASS and FAIL data distributions.
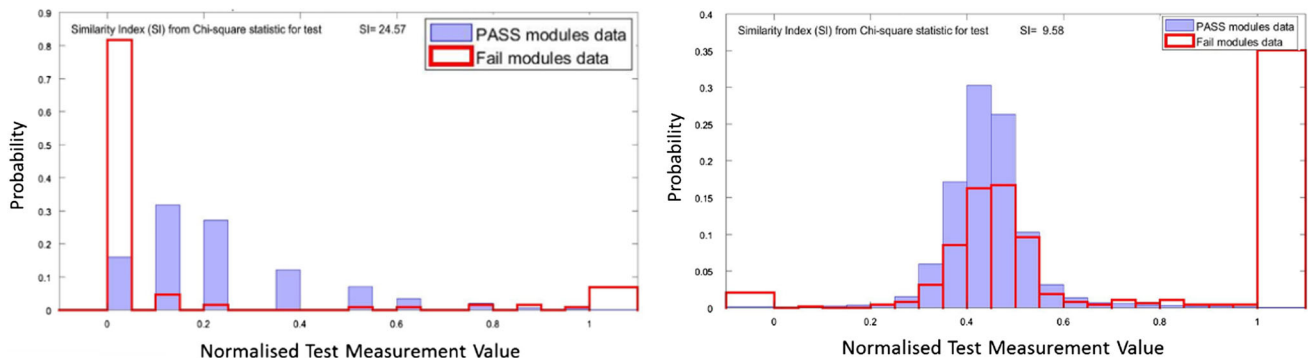
Fig. 7 Example of two qualification tests with dissimilar distributions of PASS and FAIL module data indicating tests are sensitive to pending (under subsequent sequential test) failure
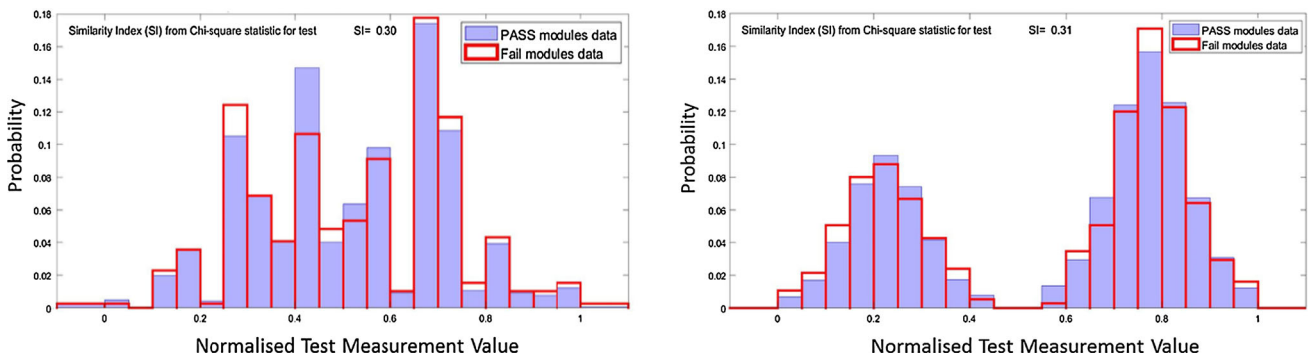
Fig. 8 Example of two qualification tests with similar distributions of PASS and FAIL module data indicating tests are not sensitive to pending (under subsequent sequential test) failure

Tests with the smallest *SI* are those tests for which distribution of PASS and FAIL data are similar. We can consider such tests as being less sensitive to detecting pending failure. The test outcomes from testing good modules and modules that fail the qualification do not differ notably and hence such test measurement data do not contain useful information to support prognostics modelling. Figure 8 details two examples of similar pairs of PASS–FAIL data distributions taken from the full set of 111 analysed qualification tests.

All 111 qualification tests with our data have been fully characterised and ranked according to their estimated similarity index but in a similar way any other application data derived from sequential tests for binary PASS–FAIL qualification can be ranked using this approach. Thus, an ordered list of tests most and least sensitive to pending failure is generated and becomes available for use in developing and demonstrating the targeted in-line prognostics capability.

Identification of redundant tests

From data analytics point of view, an individual test can be considered as a candidate for redundancy if the following two test attributes exists simultaneously:

1. No, or near zero (if acceptable), failures under the test.
2. The test outcome, in terms of distribution of the result values, follows similar distribution as the distribution of one or more preceding tests in the qualification test sequence.

From a simplistic point of view, meeting (1) above alone might be seen as a sufficient condition alone to consider/decide on potential test redundancy. If, in a large enough dataset, no modules fail that test then clearly the role of that test is somewhat less important. The challenge here is that it is very difficult to judge, in practical terms, what size of analysed data can provide sufficient confidence that indeed the probability of a module failure under that test is indeed zero or near zero. In practice, that judgement will have to accommodate the acceptable failure rate for the tested device.

Hence, we propose to add a second requirement, under (2) above, which aims at the identification of test(s) that have the same, or very similar, characteristics of the test result data. Similarity in the measured test data behaviour would imply the earlier test would detect and trigger the failed test condition, and all modules which, have passed the earlier test, will now pass those that follow and have similar test measurement characteristics. The tested physical aspect of the module with the particular test is not accounted for here. This considera-

tion is important and will need to be considered by product engineers in the light of the presented results from this data analytics.

The calculated test similarities are based on pairs of tests (test i and test j) and involve assessment of their respective PASS data distributions. This has been done for all possible pairs of tests (i, j) within the set of 111 analysed qualification tests (i.e. $i = 1, 110$ and $j = (i + 1)$, 111). The similarity evaluations use the calculated p-value in the Chi square goodness of fit test, as detailed previously in the paper.

Figure 9 details an example of identified group of four qualification tests with similarity in the test data behaviour. For each test, the respective number of failures under the test is also considered; this is to ensure identifying tests that meet both requirement (1) and (2) above. A test in a group with zero failures and preceded by another test (or other tests) in the group can potentially be considered redundant.

In-line prognostics

The main benefit of the data analytics detailed in the previous sections is the knowledge generated on the tests defined as being sensitive to pending failure. Such tests are assumed to be run first in the sequence, and would form the training data required by machine learning methods. Thus the test values will be associated with the model inputs for the constructed machine learning predictive model. Models built from such test data will be more robust, less complex in terms of model structure (as a result can be run faster) and more accurate. A prognostics model, with input data from the first k completed tests in the qualification sequence, is used the forecast the overall outcome of the qualification—it makes a prediction if the current DUT will pass all remaining tests in the test sequence or not.

Prognostics model development using support vector machine

The demonstrations in our work rely on the use of Support Vector Machine (SVM) models for binary classification (Vapnik 1995; Cristianini and Shawe-Taylor 2000; Ben-Hur et al. 2001; Hastie et al. 2008). A support vector machine is constructed from data by finding the "best hyperplane" that separates the data points into the respective two classes (in this application the two classes are Pass and Fail devices). The "best hyperplane" is defined as the one providing the largest margin of separation between the points in the two classes.

As with all machine-learning methods, a SVM model is developed from the so-called training dataset so that the unknown parameters of the chosen model structure are calculated through solving an optimisation problem that provides the smallest error between the model prediction for the classification class ($- 1$ or 1, for FAIL and PASS respectively) and the actual target outcome. As pointed out in Sect. 3, the MATLAB programming environment is used to develop and demonstrate the models reported in this paper (Matlab Release 2018).

From a mathematical point of view, if N is the number of the training data points to build the SVM classifier, x_i is the feature vector ($x_i \in R^d$, $i = 1, N$) and y_i is the associated binary outcome ($y_i \in \{-1, 1\}$), a binary SVM classifier $f(x)$ is defined so that

$$f(x_i) = \begin{cases} \geq 0 & for \ y_i = +1 \\ < 0 & for \ y_i = -1 \end{cases} \tag{6}$$

Thus, the condition for correct classification becomes: $y_i f(x_i) > 0$.

In this work, we adopt the dual version of binary SVM classifier that in its linear, dual form formulation can be expressed as (Cristianini and Shawe-Taylor 2000):

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i \left(x_i^T x \right) + b \tag{7}$$

To construct the model, the training data (x_i, y_i), $i = 1, \ldots, N$ is used, and the following optimisation problem with regards the unknown vector $\alpha \left(\alpha R^N \right)$ is solved in the process of SVM model construction:

$$\max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{j,k} \alpha_j \alpha_k y_j y_k \left(x_j^T x_k \right)$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$
$$0 \leq \alpha_i \leq C, \quad i = 1, \ldots, N \tag{8}$$

where C is the so-called penalty parameter. Larger values of C put the emphasis during training of the SVM on a stricter separation between the two binary classes. If the value of C is reduced, towards 0, this makes misclassification for the SVM model less important. The penalty parameter has been set to 1 for all models developed in this study.

In the case when the data cannot be separated by simple hyperplane using the above linear SVM model, non-linear variants of SVM classifiers can produce results that are more accurate. The models developed and reported in the paper are all non-linear SVM binary classifiers that are developed by mapping of the original feature vector space of X, R^d, onto higher dimension transformed feature space of $\Phi(x)$, R^D (d<D), where data in hand becomes linearly separable:

$$\Phi : X \rightarrow \Phi(X), R^d \rightarrow R^D \tag{9}$$
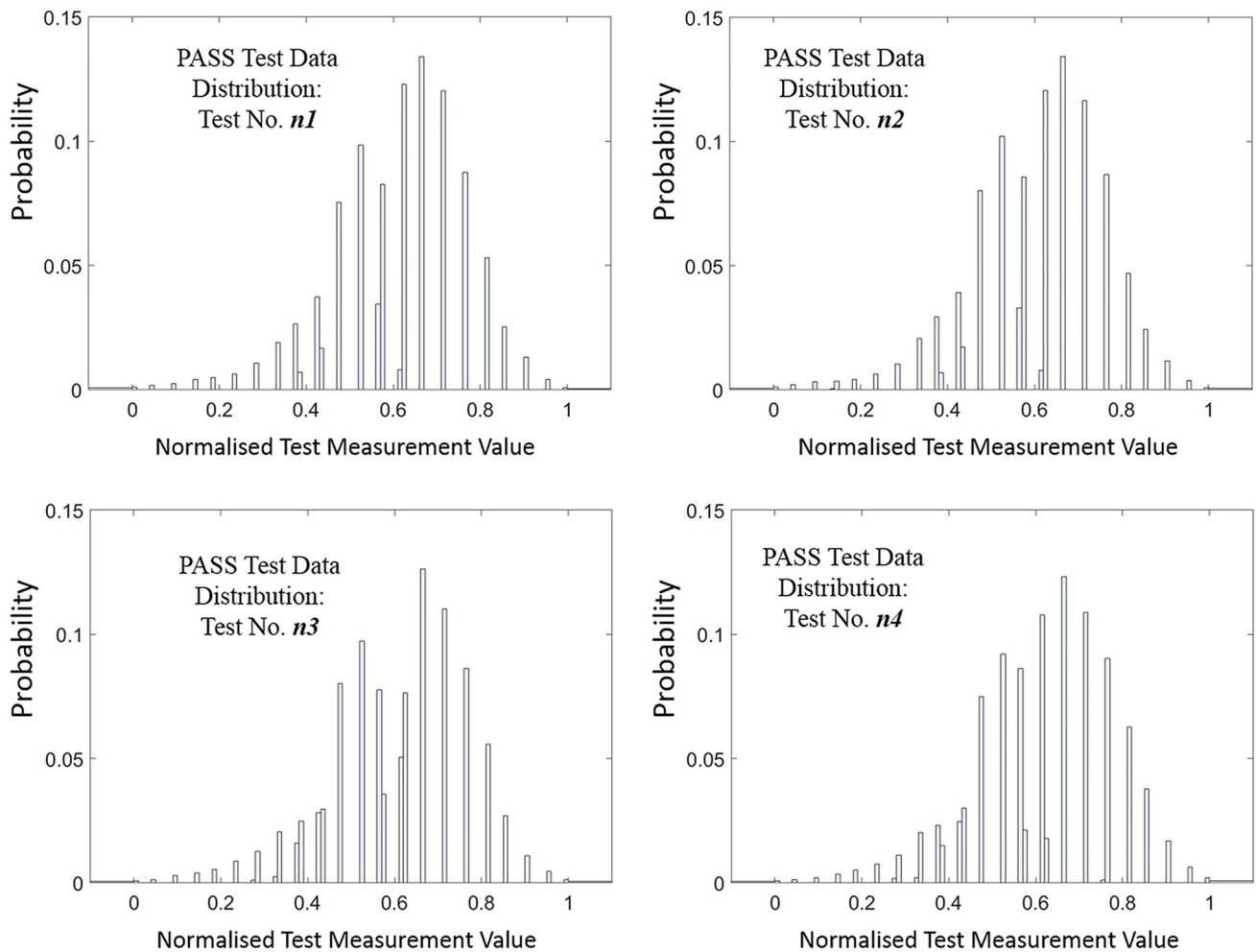
Fig. 9 Example of group of four qualification tests with similar distributions of their PASS module data

Constricting the non-linear SVM uses of the kernel trick—it is not required to explicitly define the transformation function Φ because the definitions of the linear SVM model and the optimisation problem to compute $\alpha(\alpha R^N)$ in the transformed high dimensional space R^D require only the dot product $<\Phi(X_j)^{\mathrm{T}}, \Phi(X_k)>$. Without the need to explicitly compute $\Phi(X)$, a kernel function, $G(x_j, x_k)$ is utilised in the computational process, defined as

$$G(x_j, x_k) = \langle \Phi(xj)\mathrm{T}, \Phi(xk) \rangle \tag{10}$$

The non-linear SVM model developed and demonstrated are developed using Gaussian kernel function.

$$G(x_j, x_k) = e^{-\|x_j - x_k\|^2} \tag{11}$$

The optimisation problem (8), formulated in the transformed space $\Phi(X)$ and using the kernel $G(x_j, x_k)$, is solved using the MATLAB implementation of the Sequential Minimal Optimization algorithm (Fan et al. 2005).

Two case studies are discussed. The first study evaluates the expected benefit of the proposed approach of using tests sensitive for pending failure to define and construct the training datasets. Results are used to prove the validity of the data mining approach that has been utilised with the aim to enable highly efficient ML predictive model development. The second case study provides a perspective on the implication of ML model accuracy for the actual production line if in-line prognostics is embedded and used. This assessment is given in the context of the results obtained using accessed real datasets and associate models developed.

The developed SVM models use training datasets comprising of test records for 1070 selected DUTs ($N = 1070$) and the validation data sets include 150 DUTs. The DUTs in both datasets are randomly selected, with the number of the PASS and FAIL DUTs in the datasets being equal (50:50 split in the data). This is an important requirement to ensure balanced information is provided when the model is developed and constructed as well as to ease the comparison of the performance for the deferent models being developed.

Fig. 10 Prognostics performance of two SVM models each using 20 qualification test results as input: (1) most sensitive to pending failure tests in the range of tests from 1 to 40 (left) and (2) least sensitive to pending failure tests in the range of tests from 1 to 40 (right)
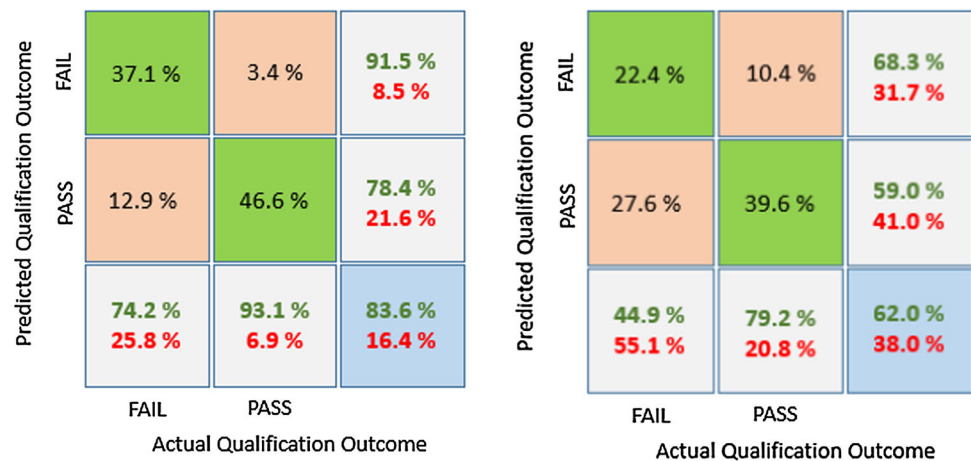
Case study 1: prognostics performance with tests sensitive to pending failure

To test the expected benefit of the definition and identification of tests sensitive to pending failure, we compare the predictive capability of two SVM models. The two models are made "equivalent" in the sense that they have identical model structure and are both using information from 20 completed tests out of the total 111 tests in the sequential test procedure. A practical way to approach this model developments and allow for model comparison in a like-to-like manner, is to assume the scenario when the DUT's have passed the first 40 tests in the sequence of tests, and test No. 40 is the current test. We use the Similarity Index (*SI*) results from the data analytics investigation and split the first 40 tests into two equal groups of tests with size 20 tests each. In the first group, we select the 20 most sensitive to pending failure tests, according to their *SI*, among the originally sequenced tests (test 1 to test 40) in the spec. Similarly, the second group represents the 20 tests identified as being the least sensitive to pending failure among the first 40 tests.

With each of the two groups of tests, in an identical manner (using exactly the same training datasets, as size and as DUT data records, and same cross-validation of generated models) we obtained SVM models with each of the two training dataset options. The inputs for each model are the 20 test results (measurements) from the respective tests, which, assuming current test completed is No. 40, are all available results at that point of the test sequence. With the training data we use the known final qualification outcome, FAIL (1) or PASS ($-$ 1), to construct the models. The SVM models make binary prediction, $-$ 1 or 1, for a given input of measurements on a device gathered from the respective 20 tests.

The predictive power and performance of the two models is evaluated using validation dataset. The model predictive accuracy for the SVM that uses 20 sensitive to pending failure tests is detailed in Fig. 10 (left) in the form of confusion

matrix plot. This is an average result obtained from large number (in this instance 100) SVM models built from different training and validation datasets (i.e. using different sets of randomly selected DUTs).

On the confusion matrix plot, the rows correspond to the predicted qualification status with the SVM (Predicted Qualification Outcome), and the columns show the actual (verified with testing) qualification status (Actual Qualification Outcome). The diagonal cells show the percentage of DUTs for which qualification is predicted correctly (actual and predicted qualification status match). The off-diagonal cells detail the % of DUTs which were not classified correctly. The far right column shows the accuracy for each predicted status and the bottom row shows the accuracy for each actual outcome. The cell in the bottom right of the plot shows the overall accuracy, in this instance 83.6%.

In a similar way, the results from the SVM model with identical complexity and same size of input information, but now based on 20 least sensitive to pending failure tests in the range of tests from 1 to 40, are summarised with Fig. 10 (right). Clear difference in the predictive capability between the two models is observed. With this prognostics model build with data from tests that are not sensitive to pending failure, the prediction accuracy for the DUTs qualification status has decreased dramatically to 62% only.

This study confirms that the approach of formulation and identification of the so-called tests sensitive to pending failure, based on similarity index attributes, is a key, integral part of the in-line prognostics strategy to smart-test execution, and offers clear improvement in the accuracy of the constructed machine learning models.

Case study 2: applied prognostics for test time/cost reduction

Case Study 2 develops SVM model using approximately 1/4 of the total number of qualification tests (actual number is

Fig. 11 Prognostics model results from SVM model using test results from 27 sensitive to pending failure tests (out of total 111 tests)

27, out of 111 tests). The tests are chosen to be tests sensitive to pending failure. In a modified and optimised qualification spec, these 27 tests will be scheduled to run first and thus will be performed before all other remaining 84 tests. Hence, this study assesses the potential benefit and the implications of using prognostics predictions for a DUT qualification in a scenario of running ¼ of the tests for our application. In this instance, time and cost reduction associated with not running the remaining ¾ of the tests (84 tests out of 111 tests) will be achieved.

Figure 11 shows the confusion matrix plot that captures the average model performance calculated from validation results gathered from 100 different validation datasets. The overall average accuracy is found to be 88.7%, with minimum and maximum overall accuracy of the SVM models from different DUT randomised training and validation datasets being 86% and 90.7% respectively.

The practical use of the model predictions, in the mode of in-line prognostics runs, will require all DUTs that receive model predictions for FAIL status to see continuing testing under the remaining tests. This is based on model-based prognostics strategy that applies:

1. Complete qualification of devices for which the model predicts FAIL status. Continuing testing under remaining tests in the qualification sequence is performed to confirm the actual qualification status of those devices.
2. Model-based qualification of the devices for which the model predicts PASS status. DUTs are qualified based on actual tests results for the first k tests (assuming model predictions performed at test k; $k = 27$ in the discussed demonstration). Remaining tests in the sequence are not executed, thus reducing qualification test time.

Let us denote the population size of the DUTs for qualification as S, with actual number of PASS devices Sp and actual number of FAIL devices Sf ($S = Sp + Sf$). In the instance of the *Case Study 2* model predictions, the number of the devices that receive FAIL status model prediction, and therefore will undergo complete testing to the end, is

$$0.797 Sf + 0.023 Sp$$

The first term accounts for the devices with actual fail status for which the model gave correct prediction FAIL (79.7%, Fig. 11, first cell in row three of the confusion matrix). The second term represents the number of devices that are actually good (Actual PASS) but were given FAIL status from the model prediction, i.e. misclassified (2.3%, Fig. 11, second cell in row three of the confusion matrix). Performing the remaining tests on these devices means that no time/cost saving is possible to achieve. However, in practice the number of these devices will represent a very small proportion from the entire population.

For this application and as an example, assume product failure in qualification is in the range of 1% ($Sf = 0.01S$ and $Sp = 0.99 * S$). This means that approximately 3% of the entire DUT population will need continuing testing beyond the 27 tests. In this instance, the balance of the DUT population (97%) would have received PASS qualification status from the model and will not be tested any further. The risk accepted with this model-based qualification approach in this instance would be that 20.3% of the DUTs that would actually FAIL the qualification, if preformed to the end, have been accepted as having PASS status based on inaccurate model classification (equivalent to up to 2 devices per 1000 tested). Thus, in the case of this production line and product, and assuming manufacturing yield of 99%, a massive 75% qualification test time/cost reduction will be achieved on 97% of the total population of DUTs if up to 2 wrongly qualified devices per 1000 DUTs can be accepted.

A prognostics model, in line with the demonstration detailed here, can be developed and employed in an identical way after the completion of different number of tests, for example once 40%, 50%, or any other percentage of the total number of tests are completed. While with larger number of completed tests the expected model accuracy may improve, the downside is that the benefit of reducing the overall qualification test time becomes lower as the model is deployed at later point in the sequence of tests.

Conclusions

This investigation aimed at the formulation and the development of a novel, computational intelligence-based approach to optimisation of qualification testing of electronic products by reducing test time and cost through off-line data analytics and imbedded in-line model-enabled prognostics. The developed approach is applicable to the most common type of

testing in the electronics industry—the multi-parameter electrical testing performed by utilising typically a large number, in the order of several hundreds, individual electrical tests. The proposed methodology and the associated models were developed, tested and validated with rigour using comprehensive datasets of real historical qualification data on an electronic module.

Mining test data with respect to failure statistic has informed on an alternative sequence for the tests execution—one that requires preforming first the tests with high failure rate—that can offer an overall qualification time reduction. Data mining for similarity using adapted Chi square test statistic/goodness-of-fit theory and the proposed similarity index-based analytics enabled a robust approach for identifying pairs and groups of tests with similar distributional behaviour in the test results data. Combining this information with failure statistic results provided the knowledge base for identifying potentially redundant tests in the qualification specification. Final decisions on redundant tests need to be taken with careful considerations and further input from test engineers considering the physical or functional aspect of the device that a test is checking.

A significant development in this research is the formulation and identification of qualification tests sensitive to pending failure using the information on uncovered data similarities. In the context of imbedded in-line prognostics capability in electronics product testing, the identification of the tests sensitive to pending failure in the qualification can support the efficient use of the data in designing training datasets for machine learning, and to build forecasting models with enhanced accuracy of the predictions for the expected qualification outcomes. This implies that further optimisation of the qualification specification is possible by moving the tests sensitive to pending failure as early as possible in the overall sequence of tests. Thus, efficient prognostics models can be adopted in the test sequence earlier than otherwise this would be possible. The time and cost benefits of the proposed test optimisation strategy were demonstrated successfully using the real qualification test datasets on the electronics module. SVM classifiers trained with only a quarter of the qualification tests, selected to be sensitive to pending failure and run first, were capable of forecasting the final outcome of the qualification with a level of accuracy close to 90%.

The proposed *smart test* methodology to optimise electrical and functional qualification test specifications of electronic devices through use of data mining and machine learning techniques, and by adopting imbedded model-based prognosis for qualification test outcomes has the potential to transform the current practices in the industry of undertaking comprehensive and time consuming testing. The major impact of this research is associated with the clear benefits, in terms of qualification test cost-time reduction, of adopting the discussed data analytics technologies, and with the presented opportunity to make this process more intelligent.

References

Ben-Hur, A., Horn, D., Siegelmann, H., & Vapnik, V. N. (2001). Support vector clustering. *Journal of Machine Learning Research, 2,* 125–137.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing, 26*(2), 213–223.

Boubezoul, A., Annanou, B., Ouladsine, M., & Paris, S. (2007). Defective wafer detection using multiple classifiers. In *Proceedings of the European control conference*, *2–5 July, Kos, Greece*, (pp. 2929–2934).

Braha, D. (Ed.). (2013). *Data mining for design and manufacturing: Methods and applications* (Vol. 179, p. 442). Berlin: Springer.

Chien, C.-F., Chang, K.-H., & Wang, W.-C. (2014). An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. *Journal of Intelligent Manufacturing, 25*(5), 961–972.

Chou, P. B., Rao, A. R., Sturzenbecker, M. C., Wu, F. Y., & Brecheret, H. B. (1997). Automatic defect classification for semiconductor manufacturing. *Machine Vision and Applications, 9,* 201–214.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing, 20*(5), 501–521.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.

Da Cunha, C., Agard, B., & Kusiak, A. (2006). Data mining for improvement of product quality. *International Journal of Production Research, 44*(18–19), 4027–4041.

Fan, R.-E., Chen, P.-H., & Lin, C.-J. (2005). Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research, 6*(2005), 1889–1918.

Godreau, V., Ritou, M., Chové, E., Furet, B., & Dumur, D. (2018). Continuous improvement of HSM process by data mining. *Journal of Intelligent Manufacturing, 1,* 1–7. https://doi.org/10.1007/s10845-018-1426-7.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning*. New York: Springer.

Kim, H. G., Han, Y. S., & Lee, J. H. (2015). Package yield enhancement using machine learning in semiconductor manufacturing. In *Proceedings of the IEEE advanced information technology, electronic and automation control conference (IEAAC), 19–20 December, Chongqing, China*, (pp. 316–320).

Kim, D., Kang, P., Cho, S., Lee, H., & Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications, 39,* 4075–4083.

Kupp, N., & Makris, Y. (2012). Integrated optimization of semiconductor manufacturing: a machine learning approach. In *Proceedings of the IEEE international test conference, 5–8 November, Anaheim, CA, USA,* (pp. 1–10).

Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research, 56*(1–2), 508–517.

MATLAB. (2018). *Statistics and machine learning toolbox.* Natick: The MathWorks Inc.

Park, S. H., Park, C. S., Kim, J. S., Kim, S. S., Baek, J. G., & An, D. (2013). Data mining approaches for packaging yield prediction in the post-fabrication process. In *Proceeding of the IEEE international congress on big data, 27 June-2 July, Santa Clara, CA, USA,* (pp. 363–368).

Pecht, M., Shibutani, T., Kang, M., Hodkiewicz, M., & Cripps, E. (2016). A fusion prognostics-based qualification test methodology for microelectronic products. *Microelectronics Reliability, 63,* 320–324.

Rokach, L., & Maimon, O. (2006). Data mining for improving the quality of manufacturing: A feature set decomposition approach. *Journal of Intelligent Manufacturing, 17*(3), 285–299.

Ruidong, D., & Chun, G. (2017). Researches and investigation on manufacturers' reliability test data of electronic parts. *Energy Procedia, 127,* 242–246.

Sohn, S. Y., & Lee, S. G. (2012). Probe test yield optimisation based on canonical correlation analysis between process control monitoring variables and probe bin variables. *Expert Systems with Applications, 39,* 4377–4382.

Stoyanov, S., Bailey, C., & Tourloukis, G. (2016). Similarity approach for reducing qualification tests of electronic components. *Microelectronics Reliability, 67,* 111–119.

Vapnik, V. N. (1995). *The nature of statistical learning theory.* New York: Springer.

Vichare, N., Zhao, P., Das, D., & Pecht, M. (2006). Electronic hardware reliability. In C. R. Spitzer (Ed.), *Avionics: Development and implementation* (pp. 414–422). Boca Raton: CRC Press.

Wang, K. (2007). Applying data mining to manufacturing: The nature and implications". *Journal of Intelligent Manufacturing, 18*(4), 487–495.

Wang, W., Azarian, M. H., & Pecht, M. (2008). Qualification for product development. In *Proceedings of the IEEE international conference on electronic packaging technology and high density packaging, 28–31 July, Shanghai, China,* (pp. 1–12).

Wuest, T., Irgens, C., & Thoben, K.-D. (2014). An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing, 25*(5), 1167–1180.