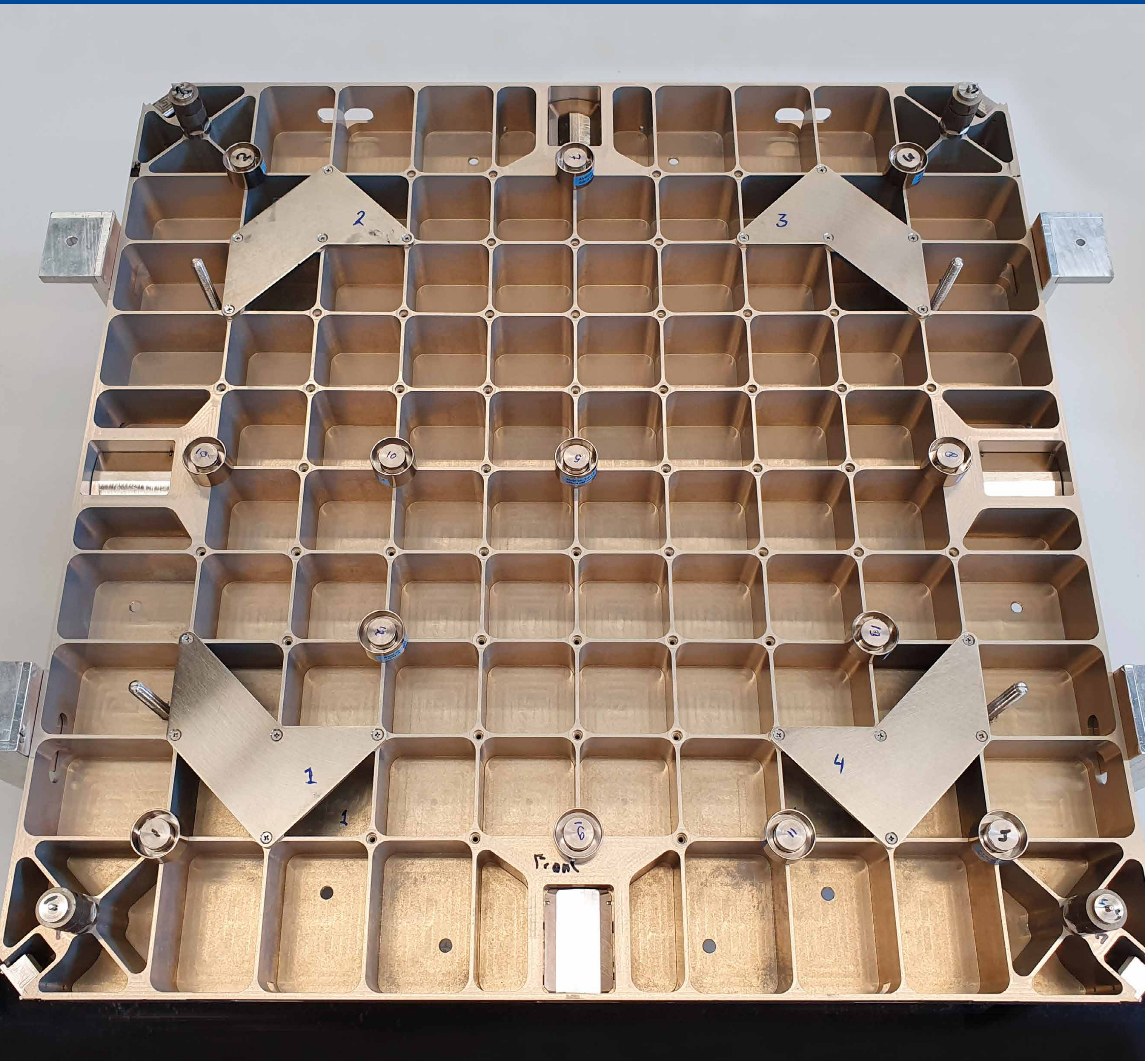
DSPE MIKRONIEK

2023 (VOL. 63) ISSUE 5

PROFESSIONAL JOURNAL ON PRECISION ENGINEERING



- THEME: SOFTWARE & MACHINE LEARNING
- FROM FAULT DIAGNOSIS AND PREDICTIVE MAINTENANCE TO CONTROL RECONFIGURATION
- DESIGN OF ROBOT SYSTEM FOR SPLICING MEDIUM-VOLTAGE POWER CABLES
- GAS BEARING WORKSHOP 2023 REPORT: DESIGN SUPPORTED BY AI
- PRECISION FAIR 2023 PREVIEW

MIKRONIEK IS A PUBLICATION OF THE DUTCH SOCIETY FOR PRECISION ENGINEERING – WWW.DSPE.NL

PUBLICATION INFORMATION

Objective
Professional journal on precision engineering and the official organ of DSPE, the Dutch Society for Precision Engineering. Mikroniek provides current information about scientific, technical and business developments in the fields of precision engineering, mechatronics and optics. The journal is read by researchers and professionals in charge of the development and realisation of advanced precision machinery.

DSPE
YOUR PRECISION PORTAL

Publisher
DSPE
Julie van Stiphout
High Tech Campus 1, 5656 AE Eindhoven
PO Box 80036, 5600 JW Eindhoven
info@dspe.nl, www.dspe.nl

Editorial board
Prof.dr.ir. Just Herder (chairman, Delft University of Technology),
Servaas Bank (VDL ETG), B.Sc.,
Maarten Dekker, M.Sc. (Philips),
Otte Haitsma, M.Sc. (Demcon),
dr.ir. Jan de Jong (University of Twente),
Erik Manders, M.Sc. (Philips Engineering Solutions),
dr.ir. Pieter Nuij (MaDyCon),
dr.ir. Ioannis Proimadis (VDL ETG),
Maurice Teuwen, M.Sc. (Janssen Precision Engineering)

Editor
Hans van Eerden, hans.vaneerden@dspe.nl

Advertising canvasser
Gerrit Kulsdom, Sales & Services
+31 (0)229 – 211 211, gerrit@salesandservices.nl

Design and realisation
Drukkerij Snep, Eindhoven
+31 (0)40 – 251 99 29, info@snep.nl

Subscription
Mikroniek is for DSPE members only.
DSPE membership is open to institutes, companies, self-employed professionals and private persons, and starts at € 80.00 (excl. VAT) per year.

Mikroniek appears six times a year.

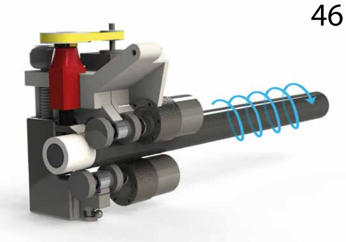
ISSN 0026-3699

The cover photo (bottom of prototype next-generation wafer stage) is courtesy of ASML and TU/e. Read the article on page 5 ff.

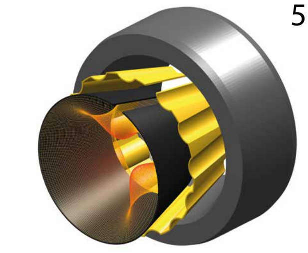
IN THIS ISSUE

THEME: SOFTWARE & MACHINE LEARNING

46



52



FEATURES

FROM EMBEDDED TO EMBODIED AI

Artificial intelligence (AI), in particular machine learning and deep learning, is on the rise in the high-tech industry, with major applications in the controlling and monitoring of complex systems. System control can be enhanced with AI to ensure better system performance, while system monitoring provides large amounts of sensor and control data that can be analysed using AI.

For example, AI can facilitate the step from corrective to predictive maintenance, with smart data analysis holding the promise of maintenance on time (thus preventing machine downtime and production losses) at lower costs (because of less damage due to wear and breakdown). This *Mikroniek* issue features several appealing examples of AI's control and maintenance potential using the concept of the much-hyped digital twin.

However, due to the focus on 'big data', machine learning in combination with software engineering has become a mainly data-related activity. But there is more to AI than data science and software engineering; it cannot be just embedded on a computer. Following the philosophical debate about AI, I plead for embodied AI, which requires the inclusion of a 'body', i.e. a machine, a robot or another instrument or physical tool. AI can only make sense if it is – through sensing and actuating – connected to the physical world.

Therefore, AI-infused systems engineering has to be complemented by other engineering disciplines – with control engineering as an overarching engineering activity – to produce optimal, more 'realistic' solutions. At universities, however, the embedded AI approach is dominant in computer-science/informatics curricula, while a holistic, multidisciplinary view of AI is lacking. Moreover, in times when energy consumption reduction is high on the agenda, we have to move away from brute-force AI modelling; for example, current deep-learning neural networks use far too many parameters, which calls for extensive computations drawing on energy-hungry data centres. We have to make more intelligent use of information from the real world, exploiting the currently underestimated power of control theory (cybernetics).

Considering the various AI approaches that are available, this means that, for example, we have to rely more on unsupervised learning. In supervised learning, training the AI model means that it is fed with input data, while the desired output is prescribed. In this approach, not all available information is used, because a real-world system produces output data as a consequence of the control it is subjected to. In unsupervised learning, however, this output can be used as additional input for the training, enabling the AI model to implicitly learn about control strategies. To state it boldly, the model develops a kind of instinct for generating the best output, just as human intelligence to a great extent relies on instinctive responses to external stimuli.

To promote a shift from embedded to embodied AI, academia has to reform its curriculum, industry has to abandon its data-centric approach of the digital twin, and government has to augment its AI policy. For example, in 2021 the Netherlands Scientific Council for Government Policy (WRR) published the report *Opgave AI. De nieuwe systeemtechnologie* (Mission AI. The new System Technology).

To wrap up, I propose adding a few recommendations concerning AI development and implementation to this report:
- to include (the connection to) the real world and, hence, more physically oriented engineering disciplines;
- to better balance the various resources, thus reducing the share of informatics as well as the volume of data used;
- to recognise the beneficial value of control theory/cybernetics and, as a consequence, gain (much needed) extra degrees of freedom in designs.

By doing this, the best inspiration source for AI comes into view – namely humans, who became intelligent without informatics!

Jan Jacobs
AI/deep learning/machine learning consultant
jan.wm.jacobs1@gmail.com

FROM FAULT DIAGNOSIS AND PREDICTIVE MAINTENANCE TO CONTROL RECONFIGURATION

Digital twins are increasingly being developed for industrial high-tech systems. In this article, we outline the exciting opportunities that digital twins have to offer for fault diagnosis, predictive maintenance, and controller reconfiguration. The proposed solutions increase economic value by minimising downtime through nonintrusive diagnostics.

KOEN CLASSENS, JEROEN VAN DE WIJDEVEN, MAURICE HEEMELS AND TOM OOMEN

Introduction

The economic value of high-tech production equipment is largely determined by its productivity, which is, in turn, heavily related to its uptime. Without maintenance, it is not a question whether a machine will fail, but rather when it will fail. Unexpected failure often results in downtime, leading to a severe loss of productivity. These unexpected breakdowns can be attributed to various factors, such as defects, the ageing of system components, and wear and tear, among others.

Ideally, critical faults are detected in an early stage and handled such that downtime is minimised. Traditionally, faults have been addressed through preventive maintenance strategies or reactive responses. In sharp contrast to these traditional approaches, predictive maintenance offers a way to further minimise equipment downtime [1,2]; see Figure 1. This is achieved by detecting faults within the equipment and precisely pinpointing their origin, a process known as fault detection and isolation (FDI) [3,4].

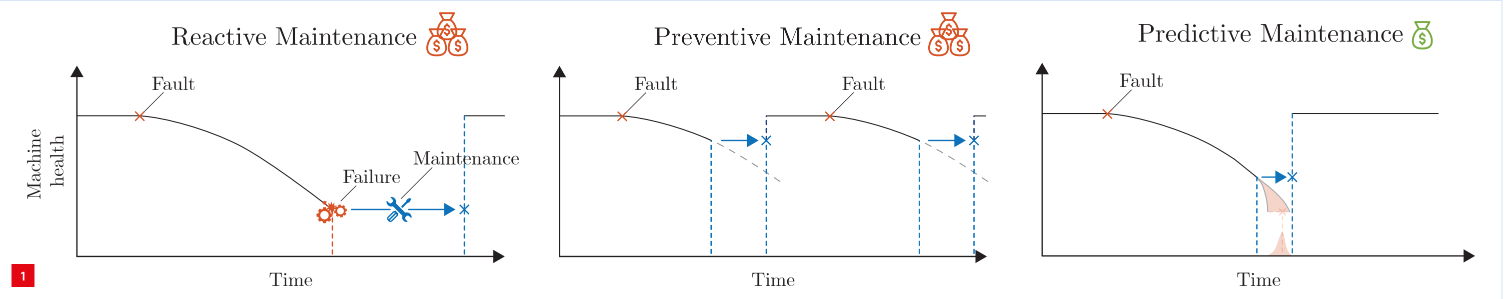
Digital twins are increasingly being developed for industrial high-tech systems [5-8]. In sharp contrast to the popularity that artificial intelligence (AI) based solutions for fault diagnosis have gained [9-11], we argue that model-based approaches should form the core foundation of digital twins for FDI in high-tech systems [12]. Typically, models of the system have been created prior to commissioning a machine. These models range in complexity from simple first-principles modelling to data-enriched finite-element modelling (FEM), or they can be identified during the system integration phase. These models, developed during the machine's design and integration phases, are at the heart of digital twins for FDI and constitute a system of interconnected systems. Despite the trend towards more AI-based solutions, we argue that the building blocks for digital twins for fault diagnosis were readily available far before the digital twin concept gained prominence in the early 2000s [13].

Surprisingly, after system integration and controller design, the developed models are often left unused. In sharp contrast, we propose to repurpose these models, since these models have a predictive power and can be harnessed in the form of a digital counterpart that is continuously informed with real-time data through already existing sensors and actuators.

AUTHORS' NOTE

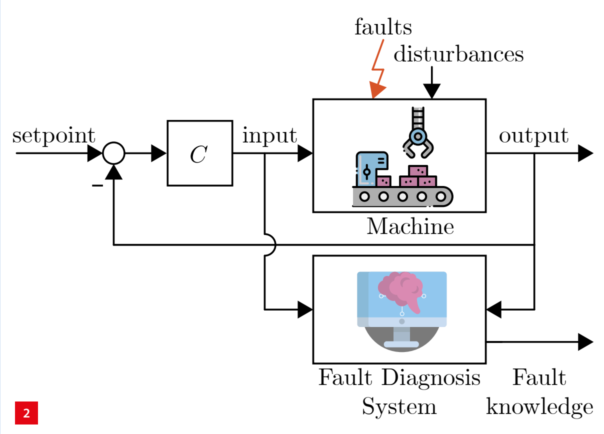
Koen Classens (Ph.D. candidate), Maurice Heemels (professor) and Tom Oomen (professor) are all associated with the Control Systems Technology group in the department of Mechanical Engineering at Eindhoven University of Technology, the Netherlands.
Jeroen van de Wijdeven is a senior researcher at ASML, Veldhoven (NL).

k.h.j.classens@tue.nl
www.tue.nl/cst
www.asml.com



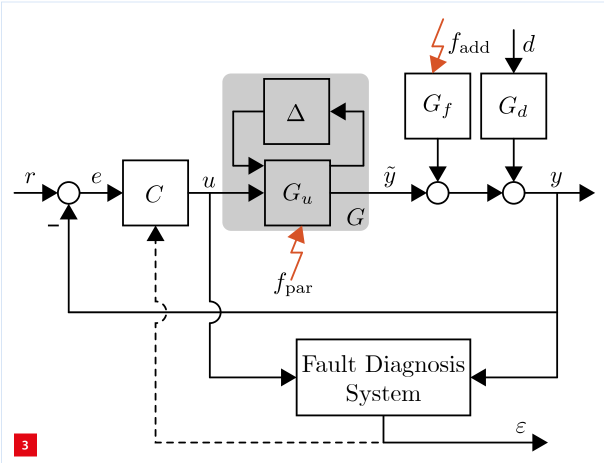
1

Illustration of three different methods to perform maintenance. Responding reactively leads to unexpected downtime and typically a long repair time. Preventive strategies lead to more interruptions than strictly necessary. Predictive maintenance is preferred as it exploits data and models to anticipate and prevent equipment failures, improving overall operational efficiency.



2

Schematic overview of a closed-loop controlled mechatronic system subjected to faults and disturbances. Data is extracted from the controlled system, which serves as input to the fault-diagnosis system that generates signals revealing knowledge about potential faults.



3

Block diagram of a closed-loop controlled system with a plant subjected to modelling uncertainty, disturbances and faults. The closed-loop controlled system is augmented with a fault-diagnosis system generating residual signals ε. The information from the fault-diagnosis system may be fed back into the controller, e.g., for reconfiguration and self-healing.

Properly designed digital counterparts have the capability to accurately predict the behaviour of the machine. Consequently, detection of an increasing mismatch between the model and the machine can indicate an upcoming failure. This yields the opportunity for scheduled instead of unscheduled service actions. As such, they provide major opportunities for enhancing uptime [12] [V1].

As the demand for higher performance increases, machines are evolving into more intricate systems. Their complexity becomes apparent through the increasing presence of actuators, sensors and components, which, in turn, generates a vast volume of data that can be harnessed for multiple purposes. Firstly, this data can be used to enhance the predictive capability of the digital counterpart. Secondly, the data can be used for monitoring and detection of anomalies. Moreover, after successful detection additional actuators and sensors might allow addressing the fault through effective self-healing. For instance, by reallocating the control effort among the remaining healthy actuators, so-called control reconfiguration.

As opposed to the opportunities arising from the abundance of actuators and sensors, this article also highlights several challenges that come with the growing complexity of systems in the context of fault diagnosis. Firstly, the increasing number of components gives rise to a greater number of potential fault scenarios, complicating the task to isolate the root cause. Secondly, for successful fault detection, it must be guaranteed that faults are distinguished from external disturbances, inherently present in any system, while simultaneously accounting for uncertainty in the system model. Thirdly, feedback controllers in mechatronic systems are designed to minimise the effects of disturbances and anomalies, making fault detection more difficult [14]. At the same time, faults in closed-loop controlled systems may be particularly hazardous as these affect machine performance and stability margins [15].

In this article, we first show the basic functionality of a model-based fault-diagnosis system and illustrate how to exploit the available models for basic FDI filter design. Subsequently, we illustrate how to give robustness guarantees and demonstrate the estimation of changing system dynamics. Finally, it is shown how to exploit fault information to enhance uptime through control reconfiguration.

Model-based fault-diagnosis system

Every fault-diagnosis system relies on data sourced from the system, such as actuator, controller and sensor data. This data can be processed offline, but also online during normal production. Fault-diagnosis systems processing this data can be designed in different ways [16-19], including data analyses, and machine-learning algorithms.

Model-based approaches acquire and process the data concurrently with the control algorithm and thus in real time, as schematically depicted in Figure 2. The output of the fault-diagnosis system is a signal or data stream that aids in identifying the presence of faults or anomalies within the system. Analysing these specifically designed signals helps diagnose and locate the specific fault in the system. So, summarising, faults that are difficult to detect directly in original data sources and/or suppressed by a feedback control loop will become visible in the output of the fault-diagnosis system.

In contrast to offline data processing, a significant advantage of online methods, such as model-based methods, is their ability to enable nearly instantaneous fault detection. One potential drawback of model-based approaches lies in their reliance on the accuracy of the underlying model. Fortunately, when it comes to mechatronic systems, precise models are readily available even before a machine is commissioned. The complexity of

AGRICULTURAL EVOLUTION WITH DEEP LEARNING

Deep learning is pivotal in modern agriculture, especially for advanced agri-robots. Using vast data, these robots discern patterns, optimising tasks from herbicide application to disease detection. However, their consistent performance hinges on a deep-learning system adept at agricultural complexities. This article explores challenges in crafting such systems, touching on economic impacts and design trade-offs. The AutoDL Platform is introduced, which is a solution for merging data & model management, task automation, and application insights.

ALBERT VAN BREEMEN

Introduction

Recent advances in deep learning are transforming agriculture (Figure 1). By processing extensive data, agri-robots recognise patterns and make precise decisions. Image recognition lets them differentiate between crops and weeds, optimising herbicide use and aiding disease prevention. They autonomously grade produce by size, colour and defects, and employ predictive analytics for improved navigation.

Agri-robots offer substantial benefits in farming. They diminish reliance on manual labour, leading to cost savings and uninterrupted operations. Their accuracy optimises seed, fertiliser, and pesticide use, boosting yields while reducing waste. Continuous monitoring facilitates early problem detection, supporting prompt interventions and informed decisions, especially in irrigation.

Agriculture's complex visuals (Figure 2), from varied crops to disease patterns, challenge traditional computer vision, which often misreads these intricacies. In contrast, deep-learning-based vision algorithms autonomously capture these details with superior precision, underpinning the advanced features of modern agri-robots.

However, deep-learning models are perpetually evolving. As fresh data surfaces and scenarios shift, models require updates to stay relevant. The ever-changing tech landscape and varying application needs drive these adjustments. Thus, a deep-learning operations platform is crucial, managing a model's lifecycle from training to monitoring, ensuring sustained peak performance. However, achieving this peak performance is not straightforward; there are numerous challenges and considerations to account for, as will be explored below.

Economics of agri-robots

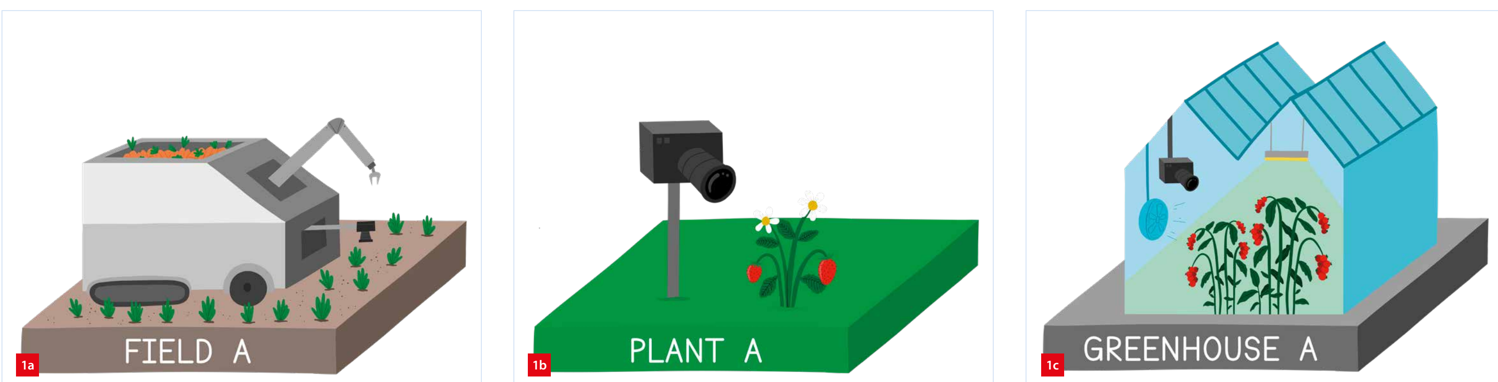
A basic economic model might help in understanding the intricacies of deep learning for agri-robots (Figure 3). To understand this model, envision a field with *M* objects ready for harvest. The agri-robot mainly consists of two parts: the visual detection system and the mechanical harvesting apparatus. In the model, if the visual system detects an object,

AUTHOR'S NOTE

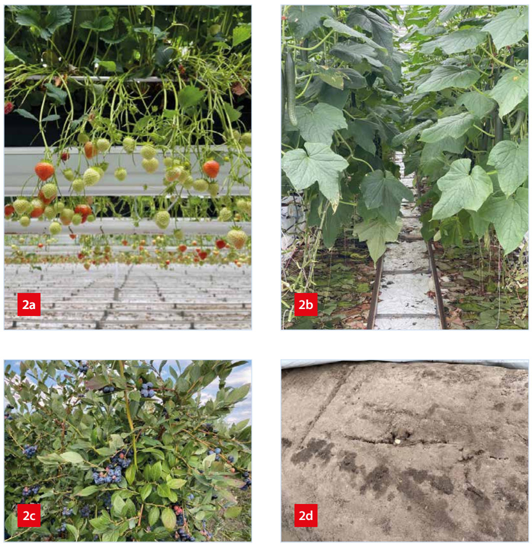
Albert van Breemen is the CEO/CTO of VBTI, an AI engineering company that develops deep-learning solutions for companies in areas such as agriculture and manufacturing.

He has over 25 years of experience in artificial intelligence, high-tech technology and innovation.

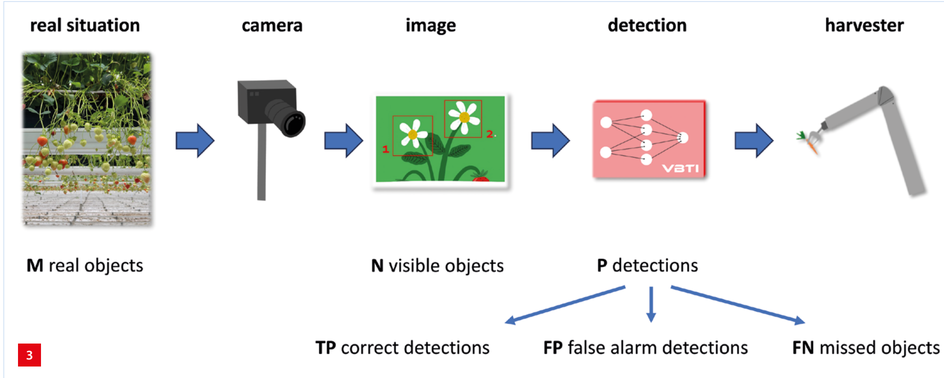
albert.van.breemen@vbti.nl
www.vbti.nl



Applications of deep learning in agriculture.
(a) Agri-robots.
(b) Intelligent monitoring.
(c) Greenhouse automation.



Typical visual agriculture scenarios.
(a) Count the number of strawberries and estimate their ripeness.
(b) Differentiate between main stem and leaf stem of cucumber plant for a leaf-cutting robot.
(c) Count number of blue berries.
(d) Detect and locate asparagus just growing above ground.



Main factors impacting system performance:
1) Object visibility v and camera properties limit the actual number of objects visible on image.
2) Detection systems are never perfect.
3) Mechanical harvester actions are for h [%] successful.

the mechanical harvester acts with h [%] 'success rate'. There is also a 'visibility factor' v [%], the ratio between the total number of objects in the field M and those visible in the image N. This factor can be enhanced by tweaking camera angles or using mechanisms to provide an unobstructed view.

Detection systems are not infallible. Given N actual objects visible in the image and P predictions, the outcomes might be:

- true positive (TP): the system correctly detects an object;
- false negative (FN): the system misses an object;
- false positive (FP): the system inaccurately detects a non-existent object.

Here, $TP + FN = N = vM$ and $TP + FP = P$.

This rudimentary economic model for agri-robots centres on two aspects. Firstly, unharvested crops result from missed detections (e.g., false negatives). Secondly, needless harvester activations, stemming from false detections where no actual object exists (e.g., false positives). The maximum revenue and nominal cost are defined by:

$$Revenue_{max} = M \cdot p$$

$$Cost_{nominal} = (M / A) \cdot c$$

Here, A [actions/hour] is the harvester throughput, c [€/hour] is the cost per machine hour, and p [€/obj] is the price per harvested object.

Depending on object visibility, harvester success rate and detection failure, the actual revenue and cost will be:

$$\begin{aligned} Revenue &= M \cdot v \cdot [TP / (FN + TP)] \cdot h \cdot p \\ &= v \cdot TPR \cdot h \cdot Revenue_{max} \\ Cost &= [(TP + FP) / A] \cdot c \\ &= [(TP + FP) / (M \cdot v)] \cdot v \cdot Cost_{nominal} \\ &= [(TP + FP) / (FN + TP)] \cdot v \cdot Cost_{nominal} \\ &= (TPR + FPR) \cdot v \cdot Cost_{nominal} \end{aligned}$$

Here, TPR is the true positive rate (or recall), and FPR is the false positive rate (or false alarm rate).

From this model, it is evident that even a slight decrease in the TPR can dent farmer profits. For example, a TPR of 95% means 5% of potential harvest is overlooked. Moreover, every inaccurate detection diminishes efficiency. For example, if for every accurate detection there is an inaccurate one, the robot's harvest efficiency drops by half. In agriculture, a mere 5% dip in yield or profit is already impactful.

This underscores the high accuracy requirements of agri-robot visual systems. In specialised tasks such as leaf cutting, where incorrect detections risk damaging a plant's main stem, precision requirements can reach a remarkable 99.9997% (which equals cutting 5 main stems in a 6-months season @ 1000 leaves/hour, 8 hours/day), underscoring the demanding nature of developing robust deep-learning solutions.

Design trade-offs

Deep-learning model development involves a complex, iterative optimisation process (Figure 4), as the efficacy of a model is deeply influenced by its architecture and the dataset it is trained on.