ICAS 2024

The Twentieth International Conference on Autonomic and Autonomous Systems

ISBN: 978-1-68558-140-4

March 10th –14th, 2024

Athens, Greece

ICAS 2024 Editors

Oliver Michler, TU-Dresden, Germany

Roy Sterritt, Ulster University, UK

ICAS 2024

Foreword

The Twentieth International Conference on Autonomic and Autonomous Systems (ICAS 2024), held between March 10 – 14, 2024, was a multi-track event covering related topics on theory and practice on systems automation, autonomous systems and autonomic computing.

The main tracks referred to the general concepts of systems automation, and methodologies and techniques for designing, implementing and deploying autonomous systems. The next tracks developed around design and deployment of context-aware networks, services and applications, and the design and management of self-behavioral networks and services. We also considered monitoring, control, and management of autonomous self-aware and context-aware systems and topics dedicated to specific autonomous entities, namely, satellite systems, nomadic code systems, mobile networks, and robots. It has been recognized that modeling (in all forms this activity is known) is the fundamental for autonomous subsystems, as both managed and management entities must communicate and understand each other. Small-scale and large-scale virtualization and model-driven architecture, as well as management challenges in such architectures are considered. Autonomic features and autonomy requires a fundamental theory behind and solid control mechanisms. These topics gave credit to specific advanced practical and theoretical aspects that allow subsystem to expose complex behavior. We aimed to expose specific advancements on theory and tool in supporting advanced autonomous systems. Domain case studies (policy, mobility, survivability, privacy, etc.) and specific technology (wireless, wireline, optical, e-commerce, banking, etc.) case studies were targeted. A special track on mobile environments was indented to cover examples and aspects from mobile systems, networks, codes, and robotics.

Pervasive services and mobile computing are emerging as the next computing paradigm in which infrastructure and services are seamlessly available anywhere, anytime, and in any format. This move to a mobile and pervasive environment raises new opportunities and demands on the underlying systems. In particular, they need to be adaptive, self-adaptive, and context-aware.

Adaptive and self-management context-aware systems are difficult to create, they must be able to understand context information and dynamically change their behavior at runtime according to the context. Context information can include the user location, his preferences, his activities, the environmental conditions and the availability of computing and communication resources. Dynamic reconfiguration of the context-aware systems can generate inconsistencies as well as integrity problems, and combinatorial explosion of possible variants of these systems with a high degree of variability can introduce great complexity.

Traditionally, user interface design is a knowledge-intensive task complying with specific domains, yet being user friendly. Besides operational requirements, design recommendations refer to standards of the application domain or corporate guidelines.

Commonly, there is a set of general user interface guidelines; the challenge is due to a need for cross-team expertise. Required knowledge differs from one application domain to another, and the core knowledge is subject to constant changes and to individual perception and skills.

Passive approaches allow designers to initiate the search for information in a knowledge-database to make accessible the design information for designers during the design process. Active approaches, e.g., constraints and critics, have been also developed and tested. These mechanisms deliver information (critics) or restrict the design space (constraints) actively, according to the rules and guidelines. Active and passive approaches are usually combined to capture a useful user interface design.

We take here the opportunity to warmly thank all the members of the ICAS 2024 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICAS 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICAS 2024 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICAS 2024 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of autonomic and autonomous systems.

We are convinced that the participants found the event useful and communications very open. We also hope that Athens provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

ICAS 2024 Chairs:

ICAS 2024 Steering Committee

Roy Sterritt, Ulster University, UK
Mark J. Balas, Texas A&M University, USA
Karsten Böhm, Fachhochschule Kufstein, Austria
Jacques Malenfant, Sorbonne Université | LIP6 Lab, France
Claudius Stern, biozoom services GmbH - Kassel | FOM University of Applied Sciences – Essen, Germany

ICAS 2024 Publicity Chairs

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

ICAS 2024

Committee

ICAS 2024 Steering Committee
Roy Sterritt, Ulster University, UK
Mark J. Balas, Texas A&M University, USA
Karsten Böhm, Fachhochschule Kufstein, Austria
Jacques Malenfant, Sorbonne Université | LIP6 Lab, France
Claudius Stern, biozoom services GmbH - Kassel | FOM University of Applied Sciences – Essen, Germany

ICAS 2024 Publicity Chairs
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

ICAS 2024 Technical Program Committee

Mubarak Abdu-Aguye, Mohamed bin Zayed University of Artificial Intelligence, UAE
Lounis Adouane, Université de Technologie de Compiègne (UTC), France
Waseem Akram, University of Calabria, Italy
Getachew Alemu, Addis Ababa University, Ethiopia
Alba Amato, Institute for High-Performance Computing and Networking (ICAR), Napoli, Italy
Arash Amini, Lehigh University, USA
Vijayan K. Asari, University of Dayton, USA
Mark Balas, Texas A&MUniversity, USA
Malek Ben Salem, Accenture Labs, USA
Julita Bermejo-Alonso, Universidad Politécnica de Madrid (UPM), Spain
Sylvain Bertrand, ONERA - The French Aerospace Lab - Centre de Palaiseau, France
Navneet Bhalla, University College London, UK
Estela Bicho, University of Minho / Centre Algoritmi / CAR group, Portugal
Karsten Böhm, Fachhochschule Kufstein, Austria
Kamel Bouzgou, Université des sciences et de la technologie Oran USTO-MB, Algeria / Université Paris Saclay - Univ. Evry, France
Estelle Bretagne, University of Picardie Jules Verne / lab MIS (modeling, information and systems), France
Ivan Buzurovic, Harvard University, Boston, USA
Paolo Campegiani, Bit4id, Italy
Alberto Castagna, Trinity College Dublin, Ireland
Wen-Chung Chang, Taipei Tech, Taiwan
Shuo Cheng, Georgia Tech, USA
Marco Claudio Campi, University of Brescia, Italy
Valérie Camps, Paul Sabatier University - IRIT, Toulouse, France
Elisa Capello, Politecnico di Torino and CNR-IEIIT, Italy
Constantin F. Caruntu, "Gheorghe Asachi" Technical University of Iasi, Romania
Samuel Cerqueira Pinto, Magna Advanced Development Center, USA
Meghan Chandarana, NASA Ames Research Center, USA

Wen-Chung Chang, National Taipei University of Technology, Taiwan
Colin Chibaya, Sol Plaatje University, South Africa
Tawfiq Chowdhury, University of Notre Dame, Indiana, USA
Stéphanie Combettes, University Paul Sabatier of Toulouse | IRIT Lab, France
Cosmin Copot, University of Antwerp, Belgium
Shreyansh Daftry, NASA Jet Propulsion Laboratory | California Institute of Technology, Pasadena, USA
Giulia De Masi, Zayed University / Rochester Institute of Technology (RIT), Dubai, UAE
Angel P. del Pobil, Jaume I University, Spain
Sotirios Diamantas, Tarleton State University | Texas A&M System, USA
Manuel J. Domínguez Morales, University of Seville, Spain
Hind Bril El Haouzi, University of Lorraine, France
Larbi Esmahi, Athabasca University, Canada
Anna Esposito, Università della Campania "Luigi Vanvitelli", Italy
Nicola Fabiano, Studio Legale Fabiano, Italy / International Institute of Informatics and Systemics (IIIS), USA
Ziyue Feng, Clemson University, USA
Hugo Ferreira, INESC TEC / Porto Polytechnic Institute, Portugal
Daniel Filipe Albuquerque, Polytechnic Institute of Viseu, Portugal
Terrence P. Fries, Indiana University of Pennsylvania, USA
Munehiro Fukuda, University of Washington Bothell, USA
Wai-keung Fung, Robert Gordon University, Aberdeen, UK
Lokesh Gagnani, LDRP-ITR | KSV University | SVKM, Gujarat, India
Zhiwei Gao, Northumbria University, UK
Jemin George, DEVCOM Army Research Laboratory (ARL), USA
Martin Giese, University Clinic of Tuebingen, Germany
Philippe Giguère, Laval University, Canada
Jordi Guitart, Universitat Politècnica de Catalunya (UPC), Spain
Maki Habib, The American University in Cairo, Egypt
Cédric Herpson, University Pierre and Marie Curie (UPMC) | LIP6, Paris, France
Randa Herzallah, University of Aston, UK
Gerold Hölzl, University of Passau, Germany
Wladyslaw Homenda, Warsaw University of Technology, Poland
Wei-Chiang Hong, Asia Eastern University of Science and Technology, Taiwan
Konstantinos Ioannidis, Information Technologies Institute - Centre for Research and Technology Hellas, Thessaloniki, Greece
Kamran Iqbal, University of Arkansas at Little Rock, USA
Minsu Jang, Electronics and Telecommunications Research Institute, South Korea
Yasushi Kambayashi, Nippon Institute of Technology, Japan
Miquel Kegeleirs, IRIDIA | Université Libre de Bruxelles, Belgium
Lial Khaluf, Independent Researcher, Germany
Hasan Ali Khattak, National University of Sciences and Technology (NUST), Islamabad, Pakistan
Igor Kotenko, SPIIRAS and ITMO University, Russia
Hak Keung Lam, King's College London, UK
Charles Lesire, ONERA/DTIS | University of Toulouse, France
Baoquan Li, Tiangong University, China
Guoyuan Li, Norwegian University of Science and Technology, Norway
Jiaoyang Li, University of Southern California, USA
Ji-Hong Li, Korea Institute of Robot and Convergence, Republic of Korea

Vadim Zhmud, Novosibirsk State Technical University, Russia
Huiyu (Joe) Zhou, University of Leicester, UK

Copyright Information

Table of Contents

Trend Analysis of Regime Change and Social Unrest with Laplace Test Statistic

Christian O. Díaz Cáez
Department of Electrical Engineering
and Computer Science
Howard University
Washington, D.C. USA
e-mail: christian.diaz@bison.howard.edu

Charles Kim
Department of Electrical Engineering
and Computer Science
Howard University
Washington, D.C. USA
email: ckim@howard.edu

***Abstract*—This research focuses on the integration of the Laplace trend statistic into a web-based collaborative platform to analyze regime changes and social unrest. The Laplace statistic, known for discerning trends, is applied to unveil any trend of changes and unrest in data sets. The platform accommodates real-time analysis and encourages inter-agency collaboration. The paper details the development of the system, including the mathematical derivation of the Laplace statistic, coding techniques, and application examples. By combining statistical methods with a versatile collaboration tool, the work aims to contribute to the evolving understanding of complex socio-political dynamics, offering a more responsive and robust approach to trend analysis and prediction of such dynamics.**

Keywords—Collaborative Platform; Fragment; Laplace Statistics; Trend Analysis.

I. INTRODUCTION

Analyzing the dynamics of regime change and social unrest requires a deeply interdisciplinary approach, blending tools from political science, sociology, and statistics [1]. Such intricacy underscores the necessity for collaboration and consensus-building across multiple organizations. Yet, these collaborations often face obstacles due to the diverse nature of data sources and varying viewpoints on the same issue. Recognizing this, we have taken our previously created collaborative web-based platform a step further [2]. In this work, we integrate trend analysis, paving the way for interactive participation and richer contributions from various stakeholders [3].

Beyond the establishment of a web-based platform, a key aspect of this research is the innovative integration of the Laplace statistic for robust trend analysis [4]. This comprehensive approach seeks to take advantage of the unique capability of the Laplace statistic to effectively distinguish between a constant and increasing rate of occurrence of complex, multifaceted events. The goal is to leverage this capability for detecting early, significant signs of regime change and social unrest, thereby enhancing the overall predictive power and accuracy of the system. The choice of the Laplace statistic is informed by its simplicity, efficiency, adaptability, and reliability, making it an excellent tool for analyzing trends in a constantly changing and unpredictable socio-political environment.

As we further unravel the complexities associated with regime shifts and social unrest, it is increasingly evident that single-source data and traditional analysis methods may not yield satisfactory results. By adopting a more integrated and interactive approach, we can more effectively utilize the vast amounts of data generated by diverse agencies, organizations, and stakeholders. Therefore, the development and implementation of a web-based platform for collaborative data collection is not merely a choice, it is a necessity. This platform must not only facilitate information gathering but also be capable of real-time analysis, providing a more dynamic and insightful understanding of regime changes and social unrest.

The volatile nature of political regimes and social movements necessitates an agile and responsive analysis system capable of quickly recognizing emerging patterns, trends, and anomalies. The application in this paper of the Laplace statistic in trend analysis adds a refined layer of sophistication to traditional approaches, offering the potential to detect nuances, subtleties, and complexities within large data sets. This statistical method, coupled with the web-based collaborative platform, represents a significant stride toward creating a predictive tool that is both responsive to the multidimensional factors influencing societal change and robust enough to accommodate variations in data quality and availability confidently.

Furthermore, the emphasis on collaboration and inter-agency communication reflects a broader paradigm shift towards collective intelligence and concerted action. In the age of information overload, the capacity to filter, interpret, and act upon relevant data is paramount. The system described in this paper offers not just a technical solution but also a conceptual framework that seeks to harmonize different perspectives, methodologies, and objectives [5]. By creating a platform that encourages shared understanding and common purpose, this work contributes to the evolving discourse on technology's role in understanding and shaping our complex socio-political landscape. Moreover, fostering a culture of inclusivity and open dialogue within the platform can foster a sense of ownership and commitment among stakeholders, leading to more effective and sustainable solutions for addressing societal challenges.

This paper is structured into several key sections: Section II delves into the development of an enhanced decision support system using trend analysis; Section III offers an in-depth examination of the mathematical derivation, interpretation, and coding techniques for trend analysis; Section IV investigates the application and testing of the Laplace statistic, complemented by examples and validation; and finally, Section V provides the conclusion to the paper.

II. ENHANCED DECISION SUPPORT SYSTEM VIA TREND ANALYSIS

The innovative construct of a web-based collaborative platform is a significant design principle that focuses on enabling cooperative interactions between stakeholders from various entities across geographical locations. This dynamic concept is a versatile tool employed in numerous applications, ranging from production, development, and marketing to service sectors, and is utilized to facilitate comprehensive collaboration and joint endeavors [6].

The imperative nature of a reliable web-based collaborative platform can hardly be overstated. It empowers resource consolidation and encourages synergistic collaboration, which allows for the identification of common patterns in data sets, thus fostering a shared understanding. This heightened demand for tailored systems motivates organizations to allocate substantial resources toward their design and development. Guided by this context, we have embarked on the development of our own platform. This platform aims to accommodate data collection from diverse sources, spanning multiple agencies and fields. Additionally, it strives to formulate global decision rules, thereby providing a comprehensive insight into a population and societal statuses [2].

In the preceding stage of our research [2], we successfully pinpointed dominant variables or attributes for threat prediction using the entropy minimum principle. Following this principle, a decision rule was formulated based on these dominant variables. This rule dictated that a specific value exceeding or falling below a determined threshold would signify a particular outcome, such as a threat. In situations where the dominant variable was a time series, monitoring its progression over time allowed real-time tracking of the rule output, especially regarding threat prediction [2]. Figure 1 illustrates the proposed web-based user interface for our platform, designed to enable users to upload their data and gain access to the resulting global decision rules.

For trend analysis, both from a qualitative and a statistical standpoint, it is noteworthy that when a variable is observed in a chronological sequence over a set duration, its random appearances denote its average time of occurrences mid-period. However, if the variable shows up more frequently towards the end of the period, it indicates a distribution with its mean located at or near the observation period end [7]. One can employ the Laplace test statistic for quantitative assessment of distribution shapes. With the Laplace trend test, distinguishing between a constant event occurrence rate and an escalating one is both straightforward and potent [8].

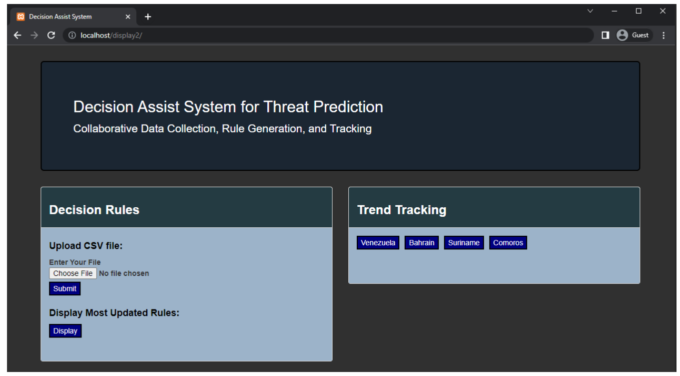


Figure 1. Proposed Webpage Interface.

III. TREND ANALYSIS

In this section, we discuss trend analysis and its applications in understanding time series data. Trend analysis helps uncover hidden patterns, fluctuations, and shifts in data over time. We will introduce the Laplace statistic, a key tool in trend analysis, and explore its significance across various fields. We will also explain how it helps identify trends and changes in event occurrences, and provide a coding implementation for real-time trend calculation and visualization.

A. Mathematical Formulations—Laplace Statistic

The Laplace statistic, also referred to as the Laplace estimator or Laplacian smoothing, is a powerful statistical tool for estimating probabilities, particularly when available data is sparse or limited. Its applications are widespread, ranging from natural language processing to machine learning. The core principle of the Laplace statistic is the adjustment of probability estimates by the addition of a small constant to both the numerator and denominator. This adjustment helps prevent zero probabilities, thereby accommodating unseen or infrequent events. This smoothing technique ensures non-zero probabilities for all possible events, enhancing the robustness and generalization capability of statistical models.

In trend analysis, the Laplace statistic emerges as an indispensable tool, shedding light on the intricacies of temporal datasets. Specifically, it quantifies the degree of smoothness or regularity within a time series, acting as a bridge between past observations and the likelihood of upcoming data points. When one embarks on the task of analyzing the Laplace statistic values plotted across a time series, it becomes possible to discern significant shifts, fluctuations, or underlying trends that might otherwise go unnoticed. This statistic is not just a mere number; it plays a pivotal role in trend analysis, offering a comprehensive lens through which we can comprehend the dynamics and recurring patterns of a given variable as it evolves over time. By harnessing the insights derived from this statistical measure, we can pinpoint both the subtle, gradual shifts and the more pronounced, abrupt changes within a dataset. This, in turn, equips decision-makers with a robust analytical foundation, ensuring decisions are data-driven, precise, and well-informed.

To exemplify the Laplace statistic utility, consider a scenario where numerous precursor events precede a final event, such as a permanent fault. The frequency of these events increases as we approach the fault. However, if similar events occur but are not precursors to the fault, their distribution over the observation period would be random. If we assume an observation period, denoted as T or $[0,T]$, with n events uniformly distributed on $[0,T]$ occurring at t_1, t_2,..., t_n. In the case of a uniform distribution over $[0,T]$, the Probability Density Function (PDF) is given by:

$$f\left(x\right)=\frac{1}{T-0}, \qquad 0\leq x\leq T. \tag{1}$$

By employing the definition of expectation, we can obtain the mean value (μ) or expectation $\mathbb{E}\left(X\right)$ for the aforementioned PDF:

$$\mu=\int_0^T x*f\left(x\right),\mathrm{d}x=\int_0^T \frac{x}{T}\,\mathrm{d}x=\frac{1}{2T}\left[x^2\right]_0^T=\frac{T}{2}. \tag{2}$$

Now, the variance $\mathbb{V}\left(X\right)$ or s with the definition of $\mathbb{V}\left(X\right)=\mathbb{E}\left(X^2\right)-\left[\mathbb{E}\left(X\right)\right]^2$ can be obtained as

$$\begin{aligned} s=&\mathbb{E}\left(X^2\right)-\left[\mathbb{E}\left(X\right)\right]^2\\ =&\int_0^T x^2 f(x)\ \mathrm{d}x-\left[\frac{T}{2}\right]^2\\ =&\frac{1}{T}\int_0^T x^2\ \mathrm{d}x-\left[\frac{T}{2}\right]^2\\ =&\frac{T^2}{12}. \end{aligned} \tag{3}$$

If we sum up all the occurrence times, the resulting sum is approximately equal to n times the mean of PDF, that is

$$\sum_{i=1}^n t_i=n*\mu=\frac{nT}{2}. \tag{4}$$

Likewise, the variance of the sum closely approximates n times the variance of the PDF, which is

$$\mathbb{V}\left(\sum_{i=1}^n t_i\right)=n*s=\frac{nT^2}{12}. \tag{5}$$

Hence, the total sum of occurrence times in a uniform distribution can be approximated by a normal distribution with the mean and the variance in (3). That can be expressed as

$$\sum_{i=1}^n t_i\approx N\left(\frac{nT}{2},\frac{nT^2}{12}\right). \tag{6}$$

If we subtract the mean occurace time ${}^{nT}/_2$ from the sum of occurrances times $\sum_{i=1}^n t_i$, the resulting difference would be zero:

$$\sum_{i=1}^n t_i-\frac{nT}{2}\approx 0. \tag{7}$$

Moreover, when we divide the aforementioned difference by the standard deviation (or the square root of the variance) of the occurrence times, the resulting equation transforms into a standard normal distribution, such as

$$\frac{\sum_{i=1}^n t_i-\frac{nT}{2}}{\sqrt{\frac{nT^2}{12}}}=\frac{\sum_{i=1}^n\left(t_i-\frac{T}{2}\right)}{T\sqrt{\frac{n}{12}}}\approx N(0,1). \tag{8}$$

The equation mentioned above represents the Laplace test statistic, which has numerous variations. In our work, we will be utilizing a specific variation of this statistic as expressed below

$$\begin{aligned} U_L=&\frac{\frac{1}{n}\sum_{i=0}^n\left(t_i-\frac{T}{2}\right)}{T\sqrt{\frac{1}{12n}}}\\ =&\frac{\mu-\frac{T}{2}}{T\sqrt{\frac{1}{12n}}}. \end{aligned} \tag{9}$$

where μ is the mean of the event occurrence times.

Equation (9) provides valuable insights into the distribution of occurrence times, enabling the identification of patterns and trends in discrete events. Nevertheless, it is crucial to acknowledge the constraints of U_L when working with smaller sample sizes. Detecting trends may pose challenges, especially when the changes in U_L occur gradually. We delve deeper into this topic in the upcoming section.

B. Interpretation for Trend Analysis

In this section, we delve deeply into the interpretation and profound significance of the statistic U_L as a valuable tool for trend analysis in discrete events. The underlying interpretation of U_L is relatively straightforward and intuitive. By examining the distribution of occurrence times, we can accurately gauge the trend exhibited by the data. If a majority of occurrence times are found to be after the midpoint ${}^T/_2$, the mean of these occurrence times (which constitutes the first term of the numerator in the U_L equation) will be larger than ${}^T/_2$. Consequently, the difference between the first and second terms of the numerator, and therefore the value of U_L, tends to be positive and often of significant magnitude. On the other hand, if a significant number of occurrence times precedes the midpoint ${}^T/_2$, the mean of the occurrence times will be smaller than ${}^T/_2$. As a result, the difference between the first and second terms of the numerator, and subsequently U_L, tends to be negative and relatively small.

As an effective indicator of trends in discrete events, U_L conveys valuable information about the direction of the trend. Its sign provides insights into the direction of the trend. A positive value indicates an upward or increasing trend, implying that the occurrence of events is predominantly happening more frequently in the latter half of the time interval T. Conversely, a negative value suggests a downward or decreasing trend, signifying that the events are occurring more frequently in the first half of the time interval.

Furthermore, we can determine the statistical significance of the observed U_L value by comparing it to the z-value associated with the standard normal distribution, which is determined by the desired confidence level. For instance, at a standard 95% confidence level, the critical z-value is

Table I. PSEUDOCODE OUTPUT FROM SAMPLE DATA.

T_i	$Sum_i\ (T_i)$	m_i	$\mu_i\ (Sum_i/m_i)$	L_i	s_i	La_i
1	1	1	1	1.73		
5	6	2	3	0.48	2.82	0.51
7	13	3	4.3	0.71	3.05	1.01
14	27	4	6.75	−0.12	5.43	−0.15
15	42	5	8.4	0.46	5.98	0.65
16	58	6	9.66	0.88	6.18	1.38
17	75	7	10.71	1.19	6.29	2.03
18	93	8	11.625	1.42	6.36	2.60
19	112	9	12.44	1.61	6.44	3.11
20	132	10	13.2	1.75	6.52	3.54
22	154	11	14	1.56	6.73	3.25
23	177	12	14.75	1.69	6.92	3.60
24	201	13	15.46	1.80	7.11	3.91
25	226	14	16.14	1.88	7.29	4.17
26	252	15	16.8	1.96	7.47	4.40

approximately 1.96. If the calculated value of U_L surpasses this threshold, it provides 95% confidence that there exists a significant increasing trend in event occurrence. Conversely, if U_L falls below -1.96, there is 95% confidence in the presence of a substantial decreasing trend. For other confidence levels, such as 90% or 99%, the corresponding threshold values for U_L are approximately 1.645 (-1.645) and 2.576 (-2.576), respectively.

It is worth noting that for situations where the sample size is small, the test for trends using U_L may encounter challenges, especially when the change in U_L is relatively slow. To address this concern, a suggested approach is to employ an adjusted Laplace test statistic that closely approximates a standard Gaussian distribution. This adjusted statistic can be obtained by multiplying the mean (μ) by the original test statistic U_L and dividing the result by the standard deviation (s) of the timed event. The resulting adjusted Laplace test statistic, denoted as U_{AL}, can be expressed as follows:

$$U_{AL} = \frac{U_L \cdot \mu}{s}. \tag{10}$$

Both the original U_L and the adjusted U_{AL} share the same critical threshold determined by the corresponding z-value, making them equally reliable tools for trend analysis, regardless of the sample size.

C. Coding for Laplace Statistic

In this section, we explore the implementation of the Laplace statistics, U_L and U_{AL}. Their calculation follows from (9), where T represents the most recent event time and n accounts for the total number of such events until time T or t_n. The derivation of U_{AL} also involves the standard deviation (s). These computations, while simple, need to be performed in real time. In essence, each new timed event leads to an update in the statistics, ensuring that the calculated Laplace statistic incorporates all past timed events, providing a comprehensive snapshot of the data.

To illustrate, we present a pseudo-code where m keeps a running total of events and $Time$ records the corresponding timestamps. Sample data demonstrating the behavior of the proposed method is showcased in Table I. A corresponding plot, Figure 2, traces the evolution of U_L and U_{AL} over time. Here, U_L reaches 1.96 at the 25th event, confirming an upward trend. Conversely, U_{AL} exhibits a faster response, crossing the 1.96 mark at the 17th event, indicating a quicker identification of the upward trend.

Building on the entropy principle discussed earlier, we draw two critical pieces of information: (1) the identification of dominant variables (or attributes) instrumental in predicting a threat and (2) a decision rule associated with these dominant variables. The rule is framed around a specific threshold, either above or below, which indicates an outcome of a threat or no threat. When dealing with a time-series dominant variable, the Laplace test statistic tracks its occurrences over time, thereby serving as a real-time threat state monitor.

In the context of binary classification, if a rule suggests that a value of 1 for a top variable leads to a Positive Transition, we then track this top variable over time. When its value is 1 at any given time, we compute all statistical values for U_L and U_{AL}. If either statistic reaches 1.96, we infer an emerging Positive Transition trend. Alternatively, if the rule indicates a Negative Transition, we track the trend accordingly. Our future plans involve applying this methodology to the top variable and rule derived from the Polity IV data [9] pertaining to regime transitions.

IV. TESTING OF LAPLACE STATISTIC

In this section, we assess the practical application of the Laplace statistic (U_L) using the Polity IV dataset. We tackle challenges like missing data and multiple entries per year per country while emphasizing trend and transition identification. You'll see how the Laplace statistic can effectively analyze real-world data for decision support.

A. Polity IV Data Set

An evaluation of decision assistance is carried out utilizing the Polity IV dataset, accessible online via the Center for Systematic Peace (CSP) [9]. This institution is committed to the innovation and enhancement of research into political violence methodologies contextualized within the ever-changing

Algorithm 1: Real-time Laplace statistic calculation

```
Input: A series of timed events Time
Output: Prints the calculated statistics for each event
Initialize tempsum to 0
for each event i in Time do
    Add the event's time to tempsum
    Calculate the cumulative sum and average time up
      to this event
    Calculate the Laplace statistic L[i] for this event
    if this is not the first event then
        Calculate the standard deviation s[i] of the time
          up to this event
        if standard deviation is not zero then
            Adjust the Laplace statistic La[i] to
              account for the standard deviation
        end
    end
    Print the statistics for this event
end
```

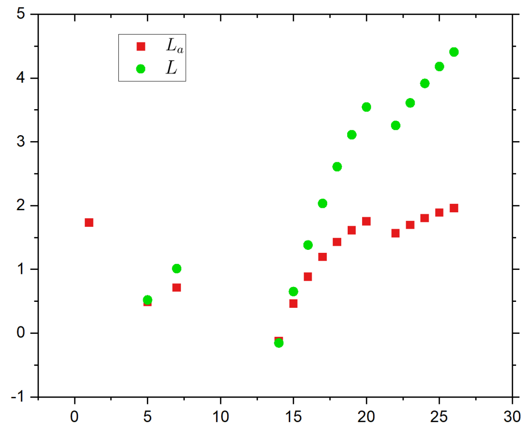


Figure 2. Laplace Statistic vs. Adjusted Laplace Statistic Over Time Axis.

framework of the global system. The CSP also provides aid for empirical research and quantitative scrutiny into human interactions and processes of socio-systemic evolution. Further, it is responsible for tracking and trend analysis of social-system performance at global, regional, and state levels, including governance aspects. The center persistently observes political behavior across all major nations, focusing on those with a population exceeding 500,000 (167 as of 2014), and reports on matters related to political violence and state failure. The date from the CSP is utilized in our research due to their commitment to open and unrestricted dissemination of their research findings [9].

Prepared by researchers at the Integrated Network for Societal Conflict Research (INSCR), the Polity IV dataset, using open-source information, is a valuable service to the broader research community. The data resources of the INSCR are rigorously cross-verified with multiple other resources, ensuring they are accurate, reliable, and complete [9]. The Polity IV project, being a treasured resource in research, has become the go-to tool for comparative and quantitative analysis, particularly useful in monitoring changes in regimes and authority and examining their effects.

The term "polity", used as a unit of analysis, refers to an organized governmental, political, or societal entity or institution, which can simply be understood as distinct patterns within an authority class. The Polity IV project amalgamates crucial information to align with the approach of examining state failures. Researchers have put significant effort into differentiating the attributes of a regime and the effectiveness of state authority, separate from the employment of organized and anti-regime armed forces that may challenge and potentially overthrow said authority.

B. Validation

In the section under discussion, we illuminate the utilization of Polity IV data for unearthing top variables and formulating rules, confronting the inherent challenges posed by the dataset structure, and probing strategies to deal with them. The Polity IV dataset encompasses the yearly political events across 194 countries, thus allowing multiple entries from different nations within the same year. Although this dataset organization facilitates the derivation of variables and rules, it simultaneously complicates trend analysis.

Our devised solution to this complication involves the formation of individual subsets for each country to ease trend analysis. A major obstacle in this approach is the presence of missing values and specific variable entries within these subsets. Two subsets that nearly fulfill our criteria are those related to Venezuela (with a country code of 101) and Bahrain (with a country code of 692).

We harness a decision-rule algorithm to identify the top variable, known as "Fragment". This variable assesses the existence of autonomous polities within the recognized state borders, capturing the degree to which the state has no effective authority over these entities [9]. The variable spans a continuous scale from 0 (no fragmentation) to 3 (serious fragmentation). It undergoes binary categorization at a threshold value of 0.0559, where values above this threshold are marked as 1, and those below it are marked as 0, signifying a Negative Transition. The starting point, or t_0, is denoted by the Polity_Begin_Year column, reflecting the inception year of the state.

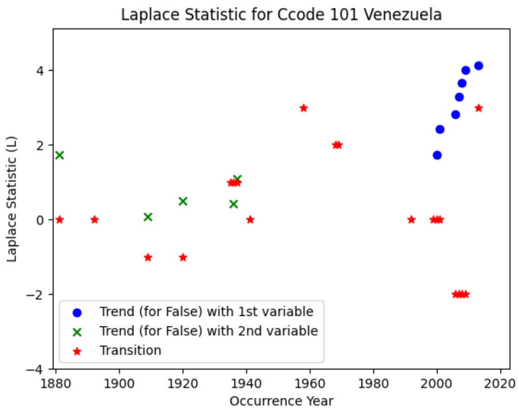
We initially applied the Laplace statistic to the Venezuelan subset. The years in Venezuela with a "Fragment" value of 0 include 2000, 2001, 2006, 2007, 2008, 2009, and 2013. The trend, as depicted in Figure 3a, reveals the first appearance of the "Fragment" value 0 at the 120^{th} year, succeeded by multiple instances of the same value. The U_L for Negative Transition transcends the 1.96 threshold and exhibits a steep upward trajectory until the final actual positive transition at the 133^{rd} year, which culminates in the rising trend. This interval of growth in U_L corresponds to the real transition phase from the 121^{st} year to the 129^{th} year. The outcome produced by the Laplace statistic code is illustrated in Figure 3c.

In the following scenario, we administer the Laplace statistic to the Bahrain subset. The years in Bahrain that feature a "Fragment" value of 0 are 2001, 2002, 2010, 2011, 2012, 2013, and 2014. The trend, highlighted in Figure 3b, discloses that at the 29^{th} year from the Polity_Begin_Year, the "Fragment" value 0 initially manifests, followed by several more such instances. The U_L for Negative Transition climbs above the 1.96 line, and subsequent to the 39^{th} year, it sharply ascends. This incrementing phase of U_L aligns with the actual negative transition period from the 40^{th} year to the 42^{nd} year. The output generated by the Laplace statistic code, based on this data, is exhibited in Figure 3d.

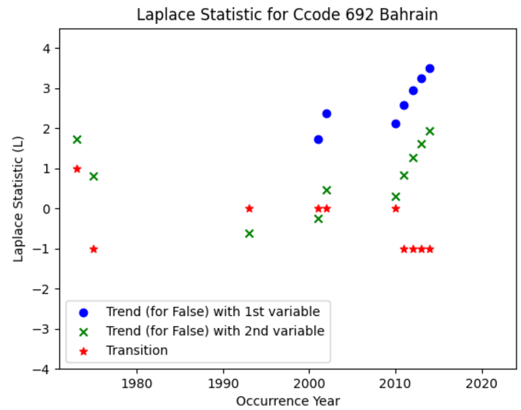
V. CONCLUSION

For expanded utilization of the Laplace trend test, we designed and tested a web-based collaboration platform to gather data from collaborators and generate decision rules on a global scale. Additionally, we thoroughly discussed the essential components of the Laplace trend test, including variable binarization, optimal variable selection using the entropy minimum principle, and the calculation of trend certainty using the maximum entropy principle. In light of evolving global

 ISBN: 978-1-68558-140-4



(a)



(b)

Laplace Statistic					
1st Variable	UL1	2nd Variable	UL2	Transition Year	Transition
0	0	0	0	1881	0
2000	1.73	1881	1.73	1892	0
2001	2.43	1909	0.08	1909	-1
2006	2.82	1920	0.5	1920	-1
2007	3.27	1936	0.43	1935	1
2008	3.65	1937	1.1	1936	1
2009	3.99			1937	1
2013	4.12			1941	0
				1958	3
				1968	2
				1969	2
				1992	0
				1999	0
				2000	0
				2001	0
				2006	-2
				2007	-2
				2008	-2
				2009	-2
				2013	3

(c)

Laplace Statistic					
1st Variable	UL1	2nd Variable	UL2	Transition Year	Transition
0	0	0	0	1973	1
2001	1.73	1973	1.73	1975	-1
2002	2.37	1975	0.82	1993	0
2010	2.11	1993	-0.62	2001	0
2011	2.58	2001	-0.24	2002	0
2012	2.94	2002	0.46	2010	0
2013	3.24	2010	0.3	2011	-1
2014	3.49	2011	0.82	2012	-1
		2012	1.25	2013	-1
		2013	1.62	2014	-1
		2014	1.93		

(d)

Figure 3. Testing results for trend analysis, showing (a) the Laplace statistic results for Venezuela, (b) the Laplace statistic results for Bahrain, (c) console output for Venezuela, and (d) console output for Bahrain.

concerns, the testing of the web server involved creating a user interface for seamless data upload, database updates, and rule adjustments using Polity IV data. By integrating the Laplace trend test into the decision-assist system with multiple data sources, our web-based platform facilitates collaborative data collection and updated decision rules, enabling global trend analysis and prediction of social unrest.

ACKNOWLEDGMENT

The authors would like to thank the financial support provided by U. S. Army through MSI STEM Research & Development Consortium with the project entitled “Irregular Warfare Decision Assist System for Determining Threat” (Award Number: D01_W911SR-14-2-0001-0015).

REFERENCES

[1] C. Ansell and A. Gash, “Collaborative governance in theory and practice,” Journal of public administration research and theory, vol. 18, no. 4, pp. 543–571, 2008.
[2] C. Díaz-Cáez and C. Kim, “A web-based data analysis platform for collaborative decision assistance,” in 23rd International Conference of Artificial Intelligence, 2021.
[3] S. Kadambe, L. G. Barajas, Y. Cho, and P. Bandyopadhyay, “System and method for signal prediction,” Mar. 1 2011, U.S. Patent 7,899,761.
[4] Y. Luo, T. Bergander, and A. B. Hamza, “Software reliability growth modelling using aweighted laplace test statistic,” in 31st Annual international Computer Software and Applications Conference (COMPSAC 2007), vol. 2. IEEE, 2007, pp. 305–312.
[5] R. Steed, T. Liu, Z. S. Wu, and A. Acquisti, “Policy impacts of statistical uncertainty and privacy,” *Science*, vol. 377, no. 6609, pp. 928–931, 2022.
[6] W. Shen, Q. Hao, and W. Li, “Computer supported collaborative design: Retrospective and perspective,” Computers in Industry, vol. 59, no. 9, pp. 855–862, 2008.
[7] C. Kim, “Identification of symptom parameters for failure anticipation by timed-event trend analysis [power distribution systems],” IEEE power engineering review, vol. 20, no. 9, pp. 48–49, 2000.
[8] H. E. Ascher and C. K. Hansen, “Spurious exponentiality observed when incorrectly fitting a distribution to nonstationary data,” IEEE transactions on reliability, vol. 47, no. 4, pp. 451–459, 1998.
[9] M. G. Marshall, “Polity IV: Regime authority characteristics and transitions datasets,” Statistical database. Vienna: Center for Systemic Peace. Available at: www.systemic peace.org/inscrdata.html (accessed July 3, 2017). Development and poverty in SSA, vol. 105, 2017.

ELaneNet: Using Lane Parameters for Better Detection of Lanes in Autonomous Driving Systems

Elikem Buertey
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada
email: ebuertey@uwaterloo.ca

Kshirasagar Naik
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada
email: snaik@uwaterloo.ca

Nitin Naik
School of Computer Science and Digital Technologies
Aston University
Birmingham, United Kingdom
email: n.naik1@aston.ac.uk

Sriram Sivaraman
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada
email: s36sivaraman@uwaterloo.ca

***Abstract*—Lane detection is crucial for an Autonomous Driving System (ADS). While traditional lane detection methods have limitations, machine learning has shown promise, though many deep learning networks struggle with variable lane detection. High Definition (HD) Maps provide comprehensive road information but are expensive and inflexible. This research proposes eLaneNet, a flexible, cost-effective, and robust lane detection system that adapts to diverse driving scenarios. By incorporating the number of lanes into the network, we demonstrate improved adaptability and potential advancements in autonomous driving technologies. We also introduce new evaluation metrics, namely, capacity, lost capacity and unsafe driving measure to assess lane detection techniques more comprehensively. We also propose evaluation of lane detection techniques by using a lane abstraction approach instead of the traditional line abstraction method. Through extensive evaluation and comparisons, we showcase the superiority of eLaneNet over LaneNet in detecting lanes. Using the TuSimple dataset, we show that eLaneNet performs better than LaneNet in detecting lanes. This research contributes to bridging the gap between ML techniques and HD maps, offering a viable solution for effective and efficient lane detection in an ADS.**

***Keywords*—Convolutional Neural Network; Enhanced LaneNet; lane detection; ELaneNet; semantic segmentation; autonomous driving; knowledge guided machine learning.**

I. INTRODUCTION

Astonishingly, 94% of road accidents are caused by human error, highlighting the potential for significant reduction in human error [1]. Addressing this issue, an Autonomous Driving System (ADS) emerges as a possible solution to decrease human error. The anticipated benefits of autonomous vehicles include crash prevention, reduced travel times, improved fuel efficiency, and parking benefits, with estimated savings of up to $2000 per year per autonomous vehicle and potentially reaching nearly $4000 when considering comprehensive crash costs [2].

Lane detection is a crucial vision problem in the context of autonomous vehicles and Advanced Driver Assistance Systems (ADAS). It involves the identification and tracking of lane markings on the road to determine the vehicle's position within its lane. In order to ensure safe and accurate travel on roads and highways, autonomous driving systems heavily depend on accurately detecting lane lines. The ultimate aim is to achieve accuracy in locating and tracking lane markings, even in challenging environmental conditions.

There is no shortage of methods, which have been suggested for lane detection [3]–[5]. Traditional methods which rely on handcrafted features and heuristics were initially proposed [6]. In challenging scenarios, including adverse weather conditions, occlusions from other vehicles, and complex urban road networks, traditional lane detection techniques often fail, highlighting the imperative for robust and adaptable solutions as emphasized in [6].

With the advent of machine learning and deep learning techniques, researchers have explored their application to the lane detection problem. However, numerous existing models have limitations when it comes to detecting an arbitrary number of lanes. This is because they are typically designed to detect a maximum of 'n' lanes, where 'n' is a specific, predefined number. Although deep learning networks such as LaneNet [6], can detect arbitrary number of lanes, there is room for improvement in this area.

Another approach to solving lane detection involves the use of HD maps. The authors in study [7] define an HD map as a map which contains all critical static properties (for example, roads, buildings, traffic lights, and road markings) of the road/environment necessary for autonomous driving, including the object that sensors cannot appropriately detect due to occlusion. HD maps, while an attractive solution,

 ISBN: 978-1-68558-140-4

are not entirely free of challenges. In particular, there is a prohibitive cost associated with producing and modifying HD maps.

In this paper, we seek to enhance lane detection accuracy by leveraging additional data while avoiding the prohibitive costs associated with HD maps. Ultimately, we aim to create a robust and adaptable lane detection system that can effectively navigate complex driving scenarios and contribute to the advancement of autonomous driving technologies.

To establish a reliable baseline for lane detection experiments, LaneNet [6] was chosen. State-of-the-art lane detection networks, such as LaneNet, take only the driving scene as input and produce an output representing the detected lane lines. However, other information on road systems is usually readily available. For example, some governments have information on road width and the number of lanes on the road on their website. We decided to explore the impact of readily available information, such as the number of lanes (NoL), on the impact of lane detection networks. We chose the NoL as the input parameter because the NoL on the road is usually constant for a long time. To do this, we modified an existing lane detection network, LaneNet, to incorporate the input image and the number of lanes on the road. We call this modified network eLaneNet. LaneNet and eLaneNet were evaluated on the TuSimple dataset to compare their performance. New metrics were also introduced to evaluate the performance of lane detection networks because of the limitations of existing metrics.

The contributions introduced in this paper are summarized as follows:

(C1) By integrating the NoL associated with a driving scene into the lane detection process, we propose eLaneNet as a new approach to address lane detection challenges.

(C2) Due to the limitations of conventional line abstracting methods, we propose that a lane abstracting method should be used to assess the performance of lane detection algorithms. In addition, we introduce capacity, lost capacity and unsafe driving measures as performance metrics since they are more specific to lane detection than general metrics, such as recall.

Experiments comparing eLaneNet to LaneNet across metrics such as capacity, lost capacity, unsafe driving measure, and accuracy consistently showed eLaneNet's superior effectiveness. This was observed in both the lane and line abstraction approaches, where entities for metric calculations were lanes and lane lines, respectively.

The structure of this paper is outlined as follows. Section II provides a review of related work.In Section III, we discuss the original LaneNet implementation and then delve into its enhanced version, eLaneNet, explaining the introduced modifications. Section IV of the paper introduces the new evaluation metrics proposed to assess the performance of lane detection systems. In Section V, we quantitatively and visually compare the performance of LaneNet and eLaneNet. Section VI concludes our work and briefly discusses possible future work.

II. Related work

In this section, we provide a comprehensive overview of pertinent literature and frequently employed datasets pertaining to Lane Detection in Autonomous Vehicles.

A. Lane Detection Methods

Various studies [8]–[11] have outlined diverse approaches for detecting and predicting lane lines.These methods can be generally classified into two categories: Traditional Methods and Deep Learning Methods.

Traditional Methods: In the era predating the rise of deep learning, lane detection relied on conventional geometric modeling techniques, such as line detection or fitting. The lane detection process usually involved four main stages: image preprocessing, feature extraction, model fitting, and lane tracking. Feature extraction encompasses the utilization of attributes, such as texture, gradients, and colors to discern essential features crucial for the identification of lane lines. Image preprocessing involved tasks such as converting colored RGB images to grayscale, reducing noise, selecting the Region of Interest (ROI), and conducting edge detection [12].

ROI selection involves vanishing point detection, perspective analysis with a projective model, and sub-sampling [13]. The idea behind using the vanishing point is that a correctly estimated vanishing point provides a strong clue about the region to localize. The authors in study [14] tackled road detection, by estimation of the vanishing point associated with the main (straight) part of the road, followed by the segmentation of the corresponding road area based on the detected vanishing point. Perspective analysis with a projective model, leverages the concept that parallel lane markings within the real-world plane converge at a vanishing point within the image plane. This approach frequently employs perspective analysis to refine the scope of detection to a precise region, which is then identified as the ROI. Through the establishment of a cohesive projection that interconnects the image plane, real-world plane, and camera plane, the process of extracting the ROI is streamlined. In study [15], a perspective projection model connects the camera and road plane, projecting lane marker edge points onto a road-space grid. The central lane line is defined by points on the grid's upper and lower edges, with each grid segment described by its offset from the lower-left point and the horizontal deviation between endpoints. In subsampling either a predefined or an adaptive region of the image is used to determine the ROI. Examples are given in [16].

Edge detection operators can be classified into Gradient and Laplacian operators, although there are additional operators that do not strictly adhere to these categories [17]. The gradient method detects edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian method uses zero crossings in the second derivative of the image to find edges. Gradient based edge detectors include Roberts, Sobel and Prewitt operators while Laplacian based edge detectors include Marrs-Hildreth edge detector. The authors in study [17] studied various edge detectors and concluded that under noisy conditions, Canny, LoG, Sobel, Prewitt, Roberts's exhibited better performance, respectively. They also concluded that Canny's edge detection algorithm has a better performance compared to the others on images.

Generally, two kinds of features exist for extracting lane lines: colors and edges [18]. Lane detection techniques can be grouped into edge-based methods, color-based methods and hybrid (edge and color) methods [13]. The Hough transform and its variants, such as, Adaptive Hough Transform and Probablistic Hough Transform are the most popular edge-based methods [13]. Steerable filter is also an edge-based technique that has been applied in many research [19] [20] [21] [22] with good results especially when road markings exhibit a clear and uniformly smooth appearance. Color-based methods have the limitation of being influenced by lighting and hence are not widely used by researchers. An example of a colour based method, HSILMD, was proposed by the authors in [23]. In HSILMD, full-color images are transformed into HSI color representation within a region of interest (ROI) to detect the road surface on the host vehicle. Using the Fuzzy c-Means algorithm, intensity distribution differences within an ROI row of pixels are recorded and clustered, enabling lane marking detection via selected intensity and saturation thresholds. Hybrid methods usually combine width, length, and location of lines with gray levels and brightness values of pixels, which improve the extraction results. An example is given in [24].

Images captured by vehicle cameras are captured in a continuous sequence. This sequential nature of image acquisition allows for an overlap between lanes detected in the current frame and those from the preceding frame. By leveraging information from both the current and previous frames, we can anticipate lane positions and track their evolution over time, enabling a more robust and accurate lane tracking process. Common trackers include Kalman filters and Particle filters [12].

Deep-learning-based methods: Deep learning lane detection methods can be grouped into: encoder-decoder CNN, Fully-Convolutional Neural (FCN) networks with optimization algorithms, CNN+RNN, and GAN model [25].

1) Encoder-decoder CNN: The encoder-decoder CNN architecture is frequently employed in semantic segmentation tasks [25]. Two examples worth discussing are LaneNet [6] and IBN-Net [26]. In the original LaneNet, which we would improve upon in this paper, an encoder-decoder network was used for binary and instance segmentation. Binary segmentation consists of segmenting the pixels into lanes and background. Instance segmentation consists of generating embeddings for lane pixels. IBN-Net improves on LaneNet by using an attention-based encoder-decoder network for lane detection. IBN-Net's encoder-decoder network also generates a binary and instance embedding. The difference between IBN-Net and LaneNet in the encoder-decoder network is that the encoder and decoder are connected by a self-attention layer.

2) FCN with optimization algorithms: In this architecture, lane detection, lane-marking identification, and vanishing-point extraction are achieved using optimization algorithms [25]. Commonly used optimization algorithms include clustering and subsampling. The architecture employs the concept of a vanishing point to guide the predictions of lane markings and road regions. VPGNet [27] leverages the vanishing point as a guiding factor and optimizes the joint prediction of lane information and road layout. This results in improved accuracy and robustness in handling complex road scenes. Deep learning methods for lane detection involving clustering is dominated by semantic segmentation algorithms. Image pixels are classified by the deep neural network, and the lane line information is extracted by clustering and other post-processing methods. An example of a deep learning method involving clustering is LaneNet.

3) CNN+RNN: Lane detection methods employing CNN+RNN architectures operate on the premise that the combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) allows for the utilization of spatial and sequential information present in road scenes respectively. Through a series of convolutional and pooling layers, CNNs are adept at extracting progressively abstract features, which are crucial for accurate lane detection. However, the analysis of road scenes necessitates the consideration of not only spatial features but also the evolution of lane configurations over time. RNNs excel at modeling sequential data by incorporating memory of prior inputs, thus capturing the dynamic nature of lane movements over frames. The unified CNN+RNN architecture therefore acquires the capacity to understand both immediate lane contexts and the evolution of the lane over time [25] [28]. This can be helpful in estimating the positions of lanes that have been obscured in the current driving scene.

4) GAN model: Given that certain lane detection methods rely on semantic segmentation, and given that Generative Adversarial Networks (GANs) are equipped to perform semantic segmentation tasks, GANs can also serve a purpose in lane detection applications. Two loss functions are used to guide the process of semantic segmentation in GANs. The generator loss function is responsible for ensuring that the GAN produces correct predictions while the discriminator loss term is more concerned with the overall image being segmented. It ensures that the individual pixel-wise predictions are consistent with each other. Thus, the discriminator evaluates the authenticity and quality of the predictions generated by the generator in a GAN. Examples of this approach are given in [29]–[31].

B. Datasets

Some common lane detection datasets include: TuSimple, BDD100K and Unsupervised LLAMAS.

1) TuSimple dataset - The TuSimple dataset was obtained from US highways and display a range of weather conditions. It consists of 358 images for validation, 2,782 images for testing and 3,626 for training, totalling 6,408 images.

2) BDD100K dataset - This dataset is drawn from more than 50,000 rides across New York and the San Francisco Bay Area city from streets, residential areas, and highways.

It contains 100K driving videos, each lasting 40 seconds. The videos are split into training (70K), validation (10K) and testing (20K) sets. The dataset is made of 720p high resolution images, with a frame rate of 30 fps and GPS/IMU recordings to preserve the driving trajectories. Ten tasks are associated with the dataset: image tagging, lane detection, drivable area segmentation, road object detection, semantic segmentation, instance segmentation, multi-object detection tracking, multi-object segmentation tracking, domain adaptation, and imitation learning.

3) Unsupervised LLAMAS dataset - Comprising of 100,042 labeled lane marker images, the unsupervised LLAMAS dataset stems from approximately 350 kilometers of recorded drives. The image labels are automatically generated, first by projecting markers into camera images, and then through further optimization to enhance label accuracy. The dataset annotations include pixel-level annotations for dashed lane markers, as well as the 3D and image space endpoints for individual markers, along with lane associations for each marker. The challenges presented within this dataset encompass a pixel-level binary segmentation problem, a segmentation problem intertwined with lane association, and a lane estimation task.

III. METHODOLOGY

In this section, we present the well-established LaneNet network. Subsequently, we introduce the enhanced version of the LaneNet network, eLaneNet.

A. LaneNet

LaneNet is structured as a two-step lane detection network as illustrated in Figure 1. Two-step lane detection methods are composed of a feature extracting step and a post-processing step [32]. LaneNet's feature extraction stage comprises the use of deep learning techniques to segment an image into two categories: binary segmentation and instance segmentation. The post-processing phase involves clustering, which groups lane pixels into clusters. Lane pixels belonging to the same lane will be in the same cluster. Finally, the fitting operation employs mathematical models to precisely define the trajectory of each lane, further enhancing the accuracy of lane boundary representation. The details of each part of LaneNet's architecture are explained below. Similar processes in both eLaneNet and LaneNet have the same number in Figure 1 and Figure 2.

Input Image (input 1) and Resize Image (process 2): In the image processing pipeline, input images are resized from their original resolution of $\alpha m \times \beta n \times c$ pixels to a reduced resolution of $(m \times n \times c)$, where $\alpha m, \beta n, m, n, c \in \mathbb{N}$. Resizing the images allows for faster computation and accommodates GPU and memory constraints.

Shared Encoder (process 3): LaneNet's shared encoder architecture is based on the ENet encoder-decoder network [33]. Two modifications to ENet's architecture was introduced in LaneNet's shared encoder. Firstly, the output of ENet was adapted to create a two-branched network, accommodating both binary segmentation and instance segmentation branches within LaneNet. Secondly, in LaneNet, only the first two stages (stages 1 and 2) of ENet's encoder are shared between the two branches, while the full ENet decoder (stages 4 and 5) serves as the backbone for each separate branch. This means that stage 3 of ENet's encoder is not used in LaneNet.

The binary segmentation branch produces a one-channel image while the instance segmentation branch produces an N-channel where N represents the embedding dimension. In this context, a k-channel image indicates that information about a specific attribute of a pixel is stored in k-dimensions. As an example, consider a color image with three channels: red, green, and blue. Each channel encodes the intensity of its respective color at each pixel, allowing for the representation of a wide spectrum of colors in the image. In the binary segmentation map of the binary segmentation branch, an output of 1 indicates that a pixel belongs to a lane instance while an output of 0 indicates that the pixel belongs to a background. In the N-channel image produced by the instance segmentation branch, an embedding dimension encodes information about which lane instance a pixel belongs to.

Segmentation Branch (process 5): The segmentation branch of the network is designed to produce a binary segmentation map, which classifies the pixels into either lane or background categories. The class weighted cross entropy loss [33] is used to account for imbalance between the lane pixels and the background pixels. As stated earlier, the output of the segmentation branch is a binary segmentation map which classifies pixels into either lane or background. Since the background pixels far exceed the lane pixels, there is an imbalance between the lane pixels and the background pixels. To address this, the class weighted cross entropy loss [33] is used to account for the imbalance between the lane pixels and the background pixels.

Embedding Branch (process 4): Embeddings produced by the embedding branch have the characteristic that lane pixels belonging to the same lane have similar embeddings while lane pixels belonging to different lanes have different embeddings. This phenomenon is achieved by using the clustering loss function in (1). In (1), the L_v term minimizes the distance between pixel embeddings belonging to the same lane. Another way to interpret this phenomenon is that the L_v term is a variance term that applies a pull force on each embedding towards the mean embedding of pixels belonging to that lane.

In order to ensure, we can distinguish between lane pixels belonging to different lanes, a distance term (L_d) is introduced. The L_d term pushes the cluster centers away from each other. In this context, a cluster center refers to the mean embedding of pixels belonging to a particular lane. Both terms are hinged. That is, the L_v term activates when an embedding is at a distance of more than δ_v from its cluster center. The pushing force, L_d between the cluster centers only activates when the centers are at a distance less than δ_d from each other. In this context, δ_d is the minimum distance allowed between cluster centers while δ_v is the maximum distance allowed between an embedding and the mean embedding of its corresponding cluster. Let K denote the number of clusters (lanes), N_k the

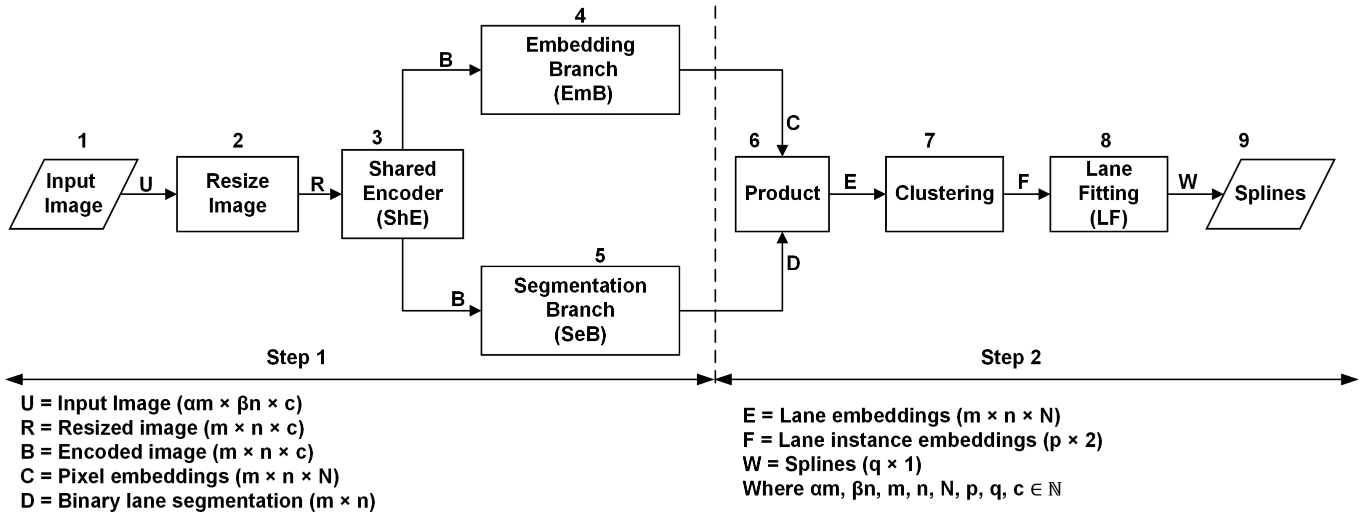


Figure 1. LaneNet architecture [6].

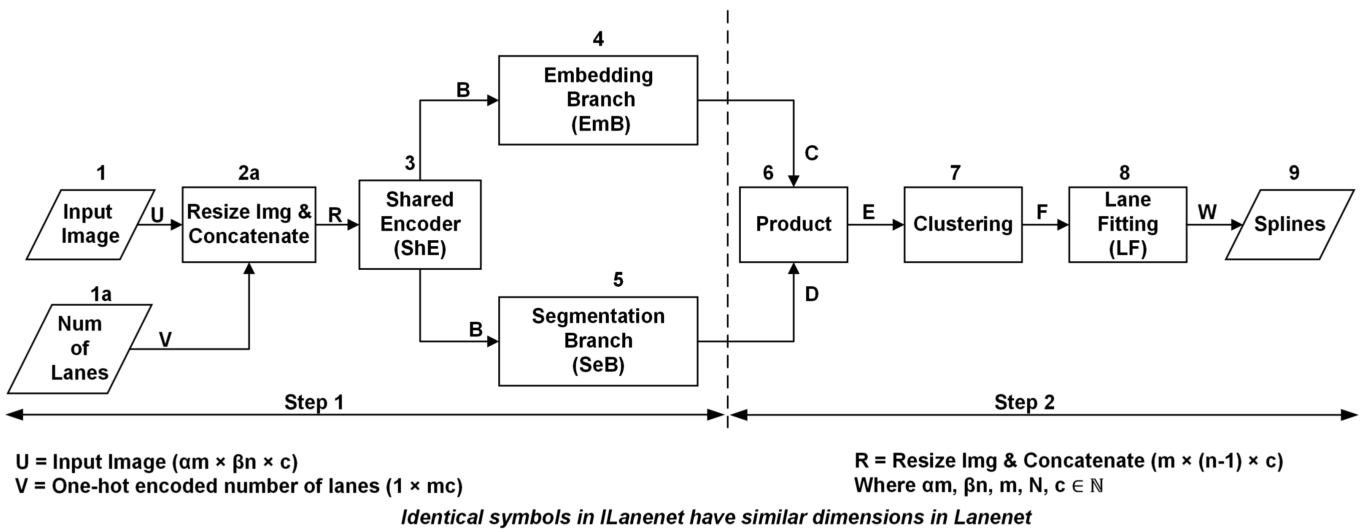


Figure 2. ELaneNet architecture.

number of elements in cluster k where $1 \leq k \leq K$, x_i is the i$^{\text{th}}$ pixel embedding in cluster k, μ_k the embedding of cluster k, $\|\cdot\|$ the L2 distance, and $[x]_+ = \max(0, x)$ the hinge, the total loss L is equal to $L_v + L_d$. The quantities L_v and L_d are defined in (1).

$$\begin{cases} L_v = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i=1}^{N_k} \left[\|\mu_k - x_i\| - \delta_v\right]_+^2 \\ L_d = \frac{1}{K(K-1)} \sum_{k_A=1}^{K} \sum_{\substack{k_B=1 \\ k_A \neq k_B}}^{K} \left[\delta_d - \|\mu_{k_A} - \mu_{k_B}\|\right]_+^2 \end{cases} \tag{1}$$

The LaneNet clustering process is performed iteratively with the loss function in (1) until the network converges. Upon convergence of the network, clusters will emerge in the embeddings of lane pixels. These clusters will exhibit a separation distance larger than δ_d from adjacent clusters, with each cluster possessing a radius smaller than δ_v.

Product (process 6) and Clustering (process 7): In the image's binary segmentation map, 0s represent background pixels, while 1s represent lane pixels. Hence, to specifically extract embeddings related to the lane pixels, we multiply the results from both the embedding and segmentation branches. This step effectively eliminates all non-lane pixels, leaving us exclusively with embeddings associated with the lanes. The subsequent application of clustering techniques helps identify distinct lane pixels corresponding to a specific lane instance. The LaneNet clustering process is executed iteratively, with the condition $\delta_d > 6\delta_v$. This condition mandates that the relationship $\delta_d > 6\delta_v$ is satisfied by the parameters δ_d and δ_v.

 ISBN: 978-1-68558-140-4

A two step approach is used when selecting a lane embedding to threshold on. First, the selected point is shifted closer to the cluster center using the mean shift algorithm. This step helps to refine the initial selection. After this shift, the thresholding process is applied, which results in the accurate identification of lane embeddings within the specified radius. This addresses the concern of accidently selecting an outlier lane embedding to threshold on.

Lane Fitting (process 8) and Splines (output 9): The process of fitting the lane in LaneNet is described as follows: First, assume we have matrix H which is the perspective transform for converting from the driving scene to bird's eye view (BEV).

Assume we have a lane pixel $\mathbf{p_i} = [x_i, y_i, 1]^T \in \mathbf{P}$ where $\mathbf{P}$ is the set of pixels belonging to a particular lane. To transform the pixel to BEV, we use H to perform the operation $\mathrm{H}\mathbf{p_i}$ where H is the perspective transformation matrix and $\mathbf{p'_i} = [x'_i, y'_i, 1]^T \in \mathbf{P'}$, the transformed pixels. After transforming the pixels to BEV, the least squares algorithm is then used to fit a low order polynomial, $f(y')$, through $\mathbf{P'}$.

To evaluate the x-value, x^*_i, of a lane at a given y-coordinate y_i, the point $\mathbf{p_i} = [-, y_i, 1]^T$ is first transformed to BEV using the expression $\mathbf{p'_i} = \mathrm{H}\mathbf{p_i} = [-, y'_i, 1]^T$. Since the current x-value in point $\mathbf{p_i}$ is irrelevant, we represent it with a -. After transformation, we use the low order polynomial with which we fitted the lane curve to evaluate the x-value of a lane at the given y position. This step is shown as follows: $x'^*_i = f(y'_i)$. We then reproject the point from BEV back to the original image space using $\mathbf{p^*_i} = \mathrm{H}^{-1}\mathbf{p'^*_i}$ where H^{-1} is the inverse perspective transformation matrix. Using this approach, we can evaluate the x-values at different y-positions.

In LaneNet, the perspective transform matrix, H, is gotten by training a neural network called HNet. In converting the image from Bird's Eye View (BEV) back to the original image space, we utilized the equation $\mathbf{p^*_i} = \mathrm{H}^{-1}\mathbf{p'^*_i}$. However, while implementing H-Net, we encountered an issue during its training phase where non-invertible matrices were outputted. This complication prevented the conversion from BEV to the original space, impeding loss calculation and leading to the suspension of network training. To address this obstacle, we chose to skip the lane fitting step altogether. Despite this omission, both LaneNet and eLaneNet exhibited satisfactory performance.

B. Improved LaneNet

ELaneNet is enhanced by a simple modification to LaneNet's architecture in step 1 and concatenating it with the resized image in step 2, while keeping the rest of the network unchanged. To augment the input image with the number of lanes (NoL), we first performed one-hot encoding on the lane count. The resulting one-hot encoded representation was then passed through a fully connected (FC) layer, which processed the lane information and produced an output ready for reshaping. This FC layer establishes connections between every input neuron and every output neuron. After reshaping, the output from the FC layer was concatenated with the original image. The resulting combined input served as the input for the LaneNet model, giving rise to a modified network known as eLaneNet.

Assume we have an initial image with dimensions $1280 \times 720 \times 3$ (αm width $\times$ βn height $\times$ c channels) and want to rescale it to a target size of $512 \times 256 \times 3$ (m width $\times$ n height $\times$ c channels). There are two different approaches with regards to LaneNet and eLaneNet.

In LaneNet, the process of rescaling is straightforward. We directly rescale the image to the target size of $512 \times 256 \times 3$.

On the other hand, the eLaneNet approach takes a slightly different route. The image is first rescaled to $512 \times 255 \times 3$ (m width $\times$ $(n-1)$ height $\times$ c channels). Additionally, the number of lanes is one-hot encoded such that the possible number of lanes ranging from 1 to $m \cdot c$ has a unique representation. To represent a lane uniquely, each position in the array corresponds to the number of lanes associated with the driving scene. Since there is only one number of lanes associated with a driving scene, one position in the array is "hot" (set to 1) while the others are "cold" (set to 0). For example, 1 lane can be encoded as $[1, 0, 0, \ldots, 0]$, 2 lanes can be encoded as $[0, 1, 0, \ldots, 0]$, 3 lanes can be encoded as $[0, 0, 1, \ldots, 0]$ and $m \cdot c$ lanes can be encoded as $[0, 0, 0, \ldots, 1]$.

This encoded lane information undergoes further processing: it passes through a fully connected layer and is reshaped into a tensor of size $512 \times 1 \times 3$ (m width $\times$ 1 height $\times$ c channels). The next step involves combining this reshaped lane information with the resized image of $512 \times 255 \times 3$. The concatenation of these two tensors results in an output tensor of size $512 \times 256 \times 3$, resembling the shape used in the LaneNet approach. This concatenated tensor, which includes both the image and the encoded lane information, serves as the input for the LaneNet network.

Figure 2 illustrates the eLaneNet architecture, while Figure 3 and Figure 4 provide a detailed depiction of the process involved in merging the inputs. The algorithm in Figure 4 first resizes the input image, then one-hot encodes the lane information, extracts lane information using a Fully Convolutional Network (FCN), reshapes the output, and finally concatenates it with the resized image to produce the output image with added lane information.

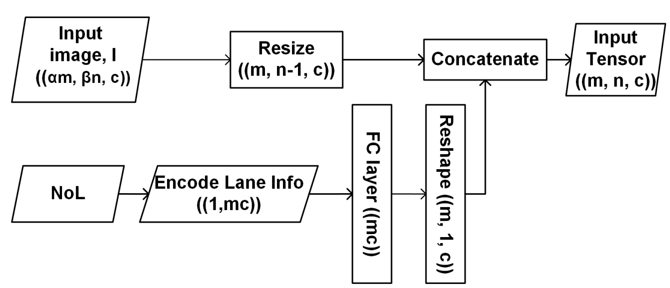


Figure 3. Concatenating image with lanes.

Algorithm 1 Add Lane Information

Input:
I: Input image (αm $\times$ βn $\times$ c)
NoL: Number of Lanes where $\mathrm{NoL} > 0$
m: Width of the target image
c: Number of channels in input image
n: Height of the target image
Output:
outImg: Image with added lane information $(\mathrm{m} \times \mathrm{n} \times \mathrm{c})$

procedure
// Resize the input image to $\mathrm{m} \times (\mathrm{n} - 1) \times \mathrm{c}$
$resizedImg \leftarrow$ ResizeImage(I, (m, n-1, c))
// One-hot encode lane information
$laneInfo \leftarrow$ EncodeLaneInformation(NoL)
// Extract lane information using FCN network
$extractedInfo \leftarrow$ FCN(laneInfo, size=mc)
// Reshape output
$laneInfo \leftarrow$ Reshape(extractedInfo, (m, 1, c))
// Concatenate lane information with the image
$outImg \leftarrow$ Concatenate(resizedImg, laneInfo)
return $outImg$
end procedure

Figure 4. Algorithm for adding lane information to an image.

IV. PERFORMANCE METRICS

Within the LaneNet framework [6], the evaluation of lane detection accuracy, represented as the average correct number of points per image, is conducted using the formula outlined in (2).

$$Accuracy = \sum_{im} \frac{C_{im}}{S_{im}} \tag{2}$$

with im denoting a driving scene in the dataset, C_{im} the total number of correctly predicted points in im and S_{im} the total number of ground-truth points in im. A correct point is defined as one where the disparity between the predicted point and the ground truth falls below a specific threshold. Equations (3) and (4) provide the formulas for computing the false positive and false negative scores, respectively.

$$\textit{False Positive Score } (FPS^l) = \frac{F^l_{pred}}{N^l_{pred}} \tag{3}$$

$$\textit{False Negative Score } (FNS^l) = \frac{M^l_{pred}}{N^l_{gt}} \tag{4}$$

with F^l_{pred} denoting the total number of falsely predicted lane lines, N^l_{pred} denoting the total number of correctly predicted lane lines, M^l_{pred} denoting the total number of missed ground-truth lane lines and N^l_{gt} denoting the total number of all ground-truth lane lines.

As described in contribution C2 of Section I, we propose the introduction of a novel performance metric referred to as the *capacity* of the lane detection system. The mathematical expression for capacity is provided in (5). To elucidate this concept, consider a scenario with two lanes on the road. When the lane detection system accurately identifies both lanes, it indicates that more ADSs can traverse the road smoothly. Essentially, this implies that the lane detection system exhibits a higher capacity by facilitating the full utilization of available lanes, thereby enhancing overall traffic flow efficiency.

On the flip side, when the system accurately identifies only one lane but fails to detect the other, it suggests the possibility of confining all vehicles to a sole lane on the road. This limitation could result in less-than-optimal utilization of the road, potentially causing traffic congestion. In these scenarios, the lane detection system is considered to have reduced capacity as it cannot efficiently utilize all available lanes, consequently hampering overall traffic efficiency. We therefore define the term "capacity" as the system's ability to detect and effectively utilize existing lane markings/lanes on the road. Conversely, lost capacity refers to the system's inability to effectively utilize existing lane markings/ lanes on the road.

We introduce a novel metric called the unsafe driving measure, alongside capacity and lost capacity considerations. The unsafe driving measure assesses the lane detection system's potential to yield inaccurate lane predictions, thereby influencing drivers to make unsafe decisions. False positives generated by the system wrongly indicate the presence of a lane. This misinformation may lead the ADS to perceive a road section as a legitimate lane, prompting an autonomous vehicle to attempt unsafe maneuvers. The corresponding expressions for capacity, lost capacity, and the unsafe driving measure are provided in (5), (6), and (7), respectively.

$$\text{Capacity}^l = \frac{TP^l}{TP^l + FN^l} \tag{5}$$

$$\text{Lost Capacity}^l = 1 - \text{Capacity}^l \tag{6}$$

$$\text{Unsafe Driving Measure}^l = \frac{FP^l}{TP^l + FN^l} \tag{7}$$

with TP^l representing the total count of accurately predicted lanes, FP^l denoting the total count of erroneously predicted lanes, and FN^l indicating the total count of missed ground-truth lanes. It's important to highlight the similarity between the expressions for capacity and the expressions for recall and false positive score, respectively.

When conceptualizing a lane as a mere line, it may lack the practical details necessary for effective use. Picture a scenario where the network outputs only a singular lane line. In this scenario, crucial information about the lane's boundaries is missing, making it difficult for a vehicle to ascertain appropriate passing areas. To overcome this limitation, we propose a more advanced lane abstraction approach.

In the proposed lane abstraction approach, lanes are considered as distinct entities as opposed to individual lane lines. This method proves beneficial by imparting a clearer understanding of the road configuration, enabling vehicles to make more informed decisions regarding lane changes and secure navigation.

To better align with the lane abstraction approach, adjustments were implemented in the following equations to account for lanes rather than lines. The expressions for the false positive score and false negative score in the lane abstraction approach are presented in Equations (8) and (9), respectively.

$$FPS^L = \frac{F^L_{pred}}{N^L_{pred}} \tag{8}$$

$$FNS^L = \frac{M^L_{pred}}{N^L_{gt}} \tag{9}$$

with F^L_{pred} as the count of incorrectly predicted lanes, N^L_{pred} as the total number of predicted lanes, M^L_{pred} as the count of missed ground-truth lanes, and N^L_{gt} as the total number of ground-truth lanes. The formulations for capacity, lost capacity, and the unsafe driving measure in the lane abstraction approach are provided in (10), (11), and (12), respectively:

$$\text{Capacity}^L = \frac{TP^L}{TP^L + FN^L} \tag{10}$$

$$\text{Lost Capacity}^L = 1 - \text{Capacity}^L \tag{11}$$

$$\text{Unsafe Driving Measure} = \frac{FP^L}{TP^L + FN^L} \tag{12}$$

with TP^L denoting the total count of accurately predicted lanes, FP^L representing the overall count of incorrectly predicted lanes, and FN^L denoting the total count of ground-truth lanes that were missed.

V. Experiments and Results

This section extensively explores experiments comparing LaneNet and eLaneNet. A thorough analysis of the results supports the claim that eLaneNet outperform LaneNet, offering valuable insights into its enhanced performance.

A. Setup

LaneNet and Improved LaneNet: The TuSimple dataset was utilized for training, and both networks were trained with an embedding dimension (N) of 4. Additionally, δ_v was set to 0.5, and δ_d was set to 3. The images underwent rescaling to 512×256. The training of the network involved using the Adam optimizer with a batch size of 32 and a learning rate of 5e-4 until convergence.

B. Performance and comparison

Quantitative Analysis: The results for the network using the line abstraction approach is given in Table I while the results for the lane abstraction approach is given in Table II.

Line Abstraction In the realm of line abstraction, the comparison between eLaneNet and LaneNet reveals that eLaneNet surpasses LaneNet across various performance metrics, establishing a subtle yet significant advantage of eLaneNet over LaneNet. The prowess of eLaneNet becomes particularly evident in its capacity to mitigate both false positives and false negatives, ultimately resulting in a higher capacity for lane detection, and an accuracy of 93.1%. This highlights eLaneNet's superior ability to accurately identify existing lane markings and effectively accommodate vehicles in all lanes, thereby minimizing instances of lost capacity on the road.

One of the noteworthy strengths of eLaneNet lies in its exceptional reduction of false positives, with a score of 13.9%, as opposed to LaneNet's 23.0%. This discrepancy underscores eLaneNet's effectiveness in minimizing the likelihood of erroneously identifying non-existent lanes as real. In the context of an ADS that leverages eLaneNet, this translates to safer and more reliable driving maneuvers. The reduced false positive rate implies a decreased risk of the system misinterpreting irrelevant features as valid lane markings, contributing to enhanced precision and reliability in autonomous navigation.

In essence, the comparative analysis demonstrates that eLaneNet not only outperforms LaneNet in terms of overall accuracy but also excels in specific aspects crucial for robust lane detection. The higher capacity and lower unsafe driving measure collectively underscore eLaneNet's advanced capabilities in identifying and interpreting lane information, making it a favorable choice for applications demanding precision and reliability in autonomous driving scenarios.

Lane Abstraction In the comparison between eLaneNet and LaneNet, the focus was on evaluating various metrics that are crucial for assessing the performance of lane detection systems. Unlike traditional approaches that assess metrics based on individual lines, this evaluation considered a more comprehensive approach by analyzing metrics at the level of entire lanes. The results, presented in Table I, shed light on the superiority of eLaneNet over LaneNet in several key aspects.

Firstly, the metric of used capacity, representing the accuracy of identifying lanes, was found to be significantly higher for eLaneNet (87.5%) compared to LaneNet (80.4%). This indicates that eLaneNet is more proficient at recognizing and delineating lanes, contributing to a more accurate representation of the road environment.

Moreover, the assessment of lost capacity, which reflects instances where the system fails to identify lanes correctly, also favored eLaneNet. The lower lost capacity of eLaneNet suggests that it experiences fewer instances of missing lanes compared to LaneNet.

In terms of safety, the metric of unsafe driving measure was introduced, and eLaneNet demonstrated a notably lower score (27.3%) compared to LaneNet (38.5%). This implies that an Autonomous Driving System (ADS) utilizing eLaneNet is

 ISBN: 978-1-68558-140-4

TABLE I. LANE ABSTRACTION

NETWORK	USED CAPACITY (RECALL)	LOST CAPACITY (FN SCORE)	UNSAFE DRIV. MEASURE (FPS SCORE)
ELaneNet	87.5 %	12.5 %	27.3 %
LaneNet	80.4 %	19.6 %	38.5 %

TABLE II. LINE ABSTRACTION

NETWORK	USED CAP (RECALL)	LOST CAPACITY (FN SCORE)	UNSAFE DRIVING MEASURE	ACC
ELaneNet	93.1 %	6.9 %	13.9 %	94.5 %
LaneNet	88.9 %	11.1 %	23.0 %	92.3 %

(a) Ground truth (b) LaneNet detects false positive (c) ELaneNet does not detect false positive

(d) Ground truth (e) LaneNet detects false positive (f) ELaneNet does not detect false positive

(g) Ground truth (h) LaneNet misses a lane (i) ELaneNet identifies one missed lane

Figure 5. Lane Detection in LaneNet and eLaneNet.

less likely to result in unsafe driving conditions compared to its counterpart LaneNet. The lower unsafe driving measure underscores the importance of accurate lane detection in enhancing the safety of autonomous vehicles.

Visual Analysis: From Figure 5 presented above, we can gain some insight into the lane detection capabilities of LaneNet and ELaneNet.

In the context of false positive detection, the ground truth images (Figures 5a and 5d) serve as the baseline, representing the actual lane markings. LaneNet's performance, as shown in Figures 5b and 5e, reveal that it tends to detect additional, false positive lane markings not present in the ground truth. This suggests that LaneNet may have a tendency to over-detect lanes in certain scenarios. Conversely, eLaneNet's results in Figures 5c and 5f demonstrate that it is more conservative in its lane detection approach. ELaneNet does not detect these false positive lane markings, which is advantageous when accuracy and avoiding false alarms are paramount.

Additionally, when considering missed lane detection, Fig. 5g represents the ground truth with all the lane markings correctly annotated. However, Figure 5h shows that LaneNet misses two of the lane markings present in the ground truth. This indicates that LaneNet may have limitations in accurately identifying all lane markings. In contrast, Figure 5i illustrates eLaneNet's ability to successfully identify one of the two missed lane marking, showcasing its strength in capturing lane markings that may be overlooked by eLaneNet.

These sample observations reveal that LaneNet exhibits a higher false positive rate and often misses lane markings, whereas eLaneNet excels in capturing missed lane markings while minimizing false positives. This makes eLaneNet a better model compared to LaneNet overall. Our LaneNet

 ISBN: 978-1-68558-140-4

results closely align with the original paper [6], especially considering the absence of accounting for conditional homography.

VI. CONCLUSION

This paper introduced an enhanced version of LaneNet designed for robust lane detection in driving scenarios. The improved architecture incorporated multiple inputs, namely, the driving scene and the number of lanes. A fully connected layer was employed to extract information from the NoL, which was then combined with the input image to create the input for LaneNet. The results demonstrated that by reducing false positives and false negatives, eLaneNet exhibited better performance compared to LaneNet. Future work aims to further enhance the model by utilizing lane count information to extrapolate missing lanes and eliminate false positives. Additionally, the effectiveness of the eLaneNet will be assessed using other datasets [34].

REFERENCES

[1] National Highway Traffic Safety Administration, “Traffic Safety Facts 2015 Data: Pedestrians,” U.S. Department of Transportation, Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115, April 2015, retrieved: December, 2023.

[2] D. J. Fagnant, K. Kockelman, “Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations,” *Transp. Res. Part A: Policy Pract.*, vol. 77, pp. 167–181, 2015. Available: https://www.sciencedirect.com/science/article/pii/S0965856415000804, retrieved: December, 2023.

[3] Z. Yang, et al., ”CANet: Curved Guide Line Network with Adaptive Decoder for Lane Detection,” in *ICASSP 2023-2023 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 1–5, 2023.

[4] J. Wang, et al., ”A keypoint-based global association network for lane detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 1392–1401, 2022.

[5] T. Zheng, et al., ”CLRNet: Cross Layer Refinement Network for Lane Detection,” in *2022 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 888–897, 2022. Available online: Available: https://doi.org/10.1109/CVPR52688.2022.00097, retrieved: December, 2023.

[6] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, L. Van Gool, “Towards End-to-End Lane Detection: an Instance Segmentation Approach,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 286–291, 2018. Available online: Available: https://doi.org/10.1109/IVS.2018.8500547, retrieved: December, 2023.

[7] Z. Bao, S. Hossain, H. Lang, X. Lin, “A review of high-definition map creation methods for autonomous driving,” *Eng. Appl. Artif. Intell.*, vol. 122, pp. 106125, 2023. Available online: Available: https://doi.org/10.1016/j.engappai.2023.106125, retrieved: December, 2023.

[8] L. Chen, Q. Li, Q. Mao, Q. Zou, “Block-constraint line scanning method for lane detection,” in *2010 IEEE Intelligent Vehicles Symposium*, pp. 89–94, 2010.

[9] A. B. Hillel, R. Lerner, D. Levi, G. Raz, “Recent progress in road and lane detection: a survey,” *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 727–745, 2014.

[10] S. Yenikaya, G. Yenikaya, E. Düven, “Keeping the vehicle on the road: A survey on on-road lane detection systems,” *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–43, 2013.

[11] Q. Li, L. Chen, M. Li, S.-L. Shaw, A. Nüchter, “A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios,” *IEEE Trans. Veh. Technol.*, vol. 63, pp. 540–555, 2014.

[12] N. B. Chetan, J. Gong, H. Zhou, D. Bi, J. Lan, L. Qie, “An Overview of Recent Progress of Lane Detection for Autonomous Driving,” in *2019 6th Int. Conf. Dependable Syst. Their Appl. (DSA)*, pp. 341–346, 2020. Available: https://doi.org/10.1109/DSA.2019.00052, retrieved: December, 2023.

[13] A. Mammeri, A. Boukerche, Z. Tang, “A real-time lane marking localization, tracking and communication system,” *Comput. Commun.*, vol. 73, pp. 132–143, 2016. Available: https://www.sciencedirect.com/science/article/pii/S0140366415003096, retrieved: December, 2023.

[14] H. Kong, J.-Y. Audibert, J. Ponce, “Vanishing point detection for road detection,” in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 96–103, 2009.

[15] Q. Li, J. Zhou, B. Li, Y. Guo, J. Xiao, “Robust Lane-Detection Method for Low-Speed Environments,” *Sensors*, vol. 18, no. 12, p. 4274, 2018. Available: https://www.mdpi.com/1424-8220/18/12/4274, retrieved: December, 2023.

[16] A. M. Shihavuddin, K. Ahmed, M. S. Munir, K. R. Ahmed, “Road boundary detection by a remote vehicle using radon transform for path map generation of an unknown area,” *Int. J. Comput. Sci. Network Secur.*, vol. 8, no. 8, pp. 64–69, 2008.

[17] G. T. Shrivakshan, C. Chandrasekar, “A comparison of various edge detection techniques used in image processing,” *Int. J. Comput. Sci. Issues (IJCSI)*, vol. 9, no. 5, p. 269, 2012.

[18] J. Niu, J. Lu, M. Xu, P. Lv, X. Zhao, “Robust Lane Detection using Two-stage Feature Extraction with Curve Fitting,” *Pattern Recognit.*, vol. 59, pp. 225-233, 2016. Available: https://www.sciencedirectz.com/science/article/pii/S0031320315004690, retrieved: December, 2023.

[19] R. K. Satzoda, M. M. Trivedi, “Drive analysis using vehicle dynamics and vision-based lane semantics,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 9–18, 2015.

[20] S. Sivaraman, M. M. Trivedi, “Integrated lane and vehicle detection, localization, and tracking: A synergistic approach,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 906–917, 2013.

[21] J. C. McCall, M. M. Trivedi, “Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, 2006.

[22] J. M. Wang, Y. C. Chung, S. L. Chang, S. W. Chen, “Lane marks detection using steerable filters,” in *Proc. 16th IPPR Conf. Comput. Vis., Graphics Image Process.*, pp. 858–865, 2003.

[23] T. Y. Sun, S. J. Tsai, V. Chan, “HSI color model based lane-marking detection,” in *2006 IEEE Intell. Transp. Syst. Conf.*, pp. 1168-1172, 2006.

[24] N. Apostoloff, A. Zelinsky, “Vision In and Out of Vehicles: Integrated Driver and Road Scene Monitoring,” *I. J. Robotic Res.*, vol. 23, pp. 513-538, 2004.

[25] Q. Zou, et al., ”Robust Lane Detection From Continuous Driving Scenes Using Deep Neural Networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 41–54, Jan. 2020.

[26] W. Li, F. Qu, J. Liu, F. Sun, Y. Wang, “A lane detection network based on IBN and attention,” *Multimed. Tools Appl.*, vol. 79, pp. 16473–16486, 2020.

[27] S. Lee, et al., ”VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition,” in *Int. Conf. Comput. Vision*, 2017, pp. 1965-1973.

[28] Y. Kortli, et al., ”Deep embedded hybrid CNN–LSTM network for lane detection on NVIDIA Jetson Xavier NX,” *Knowledge-Based Syst.*, vol. 240, p. 107941, 2022. Available: https://doi.org/10.1016/j.knosys.2021.107941, retrieved: December, 2023.

[29] Y. Liu, J. Wang, Y. Li, C. Li, W. Zhang, “Lane-GAN: A Robust Lane Detection Network for Driver Assistance System in High Speed and Complex Road Conditions,” *Micromachines*, vol. 13, no. 5, p. 716, 2022. Available: https://www.mdpi.com/2072-666X/13/5/716, retrieved: December, 2023.

[30] M. Ghafoorian, C. Nugteren, N. Baka, O. Booij, M. Hofmann, “EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection,” in *Proc. Eur. Conf. Comput. Vision (ECCV) Workshops*, 2019, pp. 256-272.

[31] Y. Zhang, Z. Lu, D. Ma, J.-H. Xue, Q. Liao, “Ripple-GAN: Lane Line Detection With Ripple Lane Line Detection Network and Wasserstein GAN,” *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1532-1542, 2021.

[32] J. Tang, S. Li, P. Liu, “A review of lane detection methods based on deep learning,” *Pattern Recognit.*, vol. 111, p. 107623, 2021. Available: https://doi.org/10.1016/j.patcog.2020.107623, retrieved: December, 2023.

[33] A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, “ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation,” *CoRR*, abs/1606.02147, June 2016.

[34] K. Wolf, et al., “Compensation mechanism in tumor cell migration: mesenchymal–amoeboid transition after blocking of pericellular proteolysis,” *J. Cell Biol.*, vol. 160, pp. 267–277, 2003.

Taking Autonomous Mobile Robots and Humans Into the Industrial Metaverse: An Empirical Study in the Automotive Industry

Taru Hakanen, Jaakko Karjalainen and Petri Tikka
Cognitive Production Industry
VTT Technical Research Centre of Finland
Tampere, Finland
email: taru.hakanen@vtt.fi

Marko Lepola and Markus Heikkilä
Telia Finland Ltd
Helsinki, Finland
email: marko.x.lepola@teliacompany.com

***Abstract*—This paper provides empirical insights into a new emerging robot-related trend, in which Autonomous Mobile Robots (AMRs) are used for building an industrial metaverse. The aims of this study were to a) develop a technical solution for AMR-enabled metaverse creation, b) to test and demonstrate the solution in a real industrial case, and c) to identify future "factory metaverse" concepts within the automotive industry. This paper presents the new solution combining an AMR-enabled digital shadow (point cloud) creation, transfer of the point cloud to an eXtended Reality (XR) platform, and data visualisation, audio and multi-user interaction in the virtual factory metaverse. The solution was tested and demonstrated in Valmet Automotive's Innovation Center in Finland. Potential future use cases and concepts were identified through a qualitative case study (n=36). The solution enables remote and multi-user situational awareness and collaboration. Potential use cases for the factory metaverse include, e.g., factory cell and work planning, remote technical support, and safety trainings.**

***Keywords*–Autonomous mobile robot; industrial metaverse; extended reality; virtual reality; digital twin.**

I. INTRODUCTION

The concept of the metaverse has been widely discussed especially since 2021, when Facebook changed their company name to Meta. Originally, the metaverse concept was mentioned as early as 1992 by Neal Stephenson in a science fiction novel called Snow Crash [1]. According to the author, the metaverse was regarded as "a three-dimensional virtual space that uses the metaphor of the real world, and where humans, as programmable avatars, interact with each other and software agents". This description still addresses the central elements of the metaverse concept quite effectively.

In the research literature, there is a lack of a comprehensive definition of the metaverse. It is defined, for instance, as "a three-dimensional online environment in which users represented by avatars interact with each other in virtual spaces decoupled from the real physical world" [2]. The metaverse involves a set of technologies such as smart devices, avatars, Artificial Intelligence (AI), Virtual Reality (VR), Augmented Reality (AR), XR, digital twins, blockchains, and robotics [3]-[5].

Recently, the metaverse has captured the attention of both academia and industry practitioners [6]. So far, the focus has predominantly been on consumer applications although the largest business potential is seen in industrial metaverse applications [7][8]. Wider application of the metaverse is still in its earliest stages and especially so in the industrial context. This study aims to provide empirical insights into metaverse applications and potential future concepts focusing specifically on the industrial B2B context. The underlying goal is to make industrial work in the manufacturing sector more efficient, safe, and motivating.

Within the industrial and operations management domain, the industrial metaverse can enable virtual employee training and collaboration, virtual prototyping, and virtual-showroom displays for products [9]. The metaverse may have many implications on product development, manufacturing, and customer relationships [3]. As COVID-19 enhanced teleworking globally, the metaverse can develop remote work and collaboration to a whole new level [10]. Arguably, the metaverse is a growing topic in production and operation research. However, more research is called for [11][12] and there is a lack of empirical studies regarding the ways industrial companies can adopt the metaverse into their operations [13].

VTT Technical Research Centre of Finland, in collaboration with 10 industrial partners, identified robotics-aided metaverse as one of the most promising industrial metaverse applications [14]. AMRs and Automated Guided Vehicles (AGVs) are nowadays commonly used for, e.g., internal logistics tasks [15]. AMRs equipped with laser scanners have the capability to sense their environments and form a 3D model (point cloud) of its surroundings for navigation and safety [15]. A link to the metaverse can thus also be built from autonomous digital shadow creation to a wider metaverse application including extended reality (XR) solutions.

The aim of this paper is to increase understanding of a robotics-assisted industrial metaverse and potential future concepts in the manufacturing sector. More specifically, the

 ISBN: 978-1-68558-140-4

aim of this study was to a) develop a technical solution for AMR-enabled metaverse creation, b) to test and demonstrate the solution in a real industrial case, and c) to identify future "factory metaverse" concepts within the automotive industry. The study results present the developed new metaverse solution, its features, and future autonomous factory metaverse concepts. A qualitative case study (n=36) was conducted in Valmet Automotive to innovate and identify potential use cases and future factory metaverse concepts.

The novelty value of this research derives from the creation of a new industrial metaverse application combining AMRs and XR, and the identification of its potential use cases in real-life industrial operations. The study is among the first empirical metaverse studies within the production and operations management domain and it provides a concrete example of the ways the industrial metaverse can be realised in the future.

This paper is organised as follows: Section 1 presents an introduction to the paper. In Section 2, theoretical background and literature synthesis are provided. Section 3 outlines the methodology of the study. Section 4 presents the study results: a) description of the new factory metaverse solution and b) identified new use cases and concepts. In Section 5, study contribution, limitations, and potential future research avenues are discussed.

II. THEORY

This Section outlines the theoretical background concerning industrial metaverse, digital twins, AMRs, and XR solutions. Finally, the Section synthesizes the literature review and research needs.

A. *Industrial metaverse and digital twins*

The industrial metaverse "combines physical-digital fusion and human augmentation for industrial applications and contains digital representations of physical industrial environments, systems, assets and spaces that people can control, communicate, and interact with" [8]. Potential applications in the production and operations management domain include, e.g., simulation and testing of different scenarios at the design stage [3]. The metaverse offers multiple opportunities for production systems, training, and agile virtual prototyping [9]. The metaverse enables people to join the design at any time from any place and communicate to each in a common digital environment. Thus, it enhances the trend of remote work and operations. The metaverse allows companies to simulate and optimise factory planning and complex industrial processes the way, for instance, BMW has done [16].

Metaverse is a technology to support the well-established frameworks and initiatives in the manufacturing sector such as Industry 4.0 and 5.0 [17]. Both streams identify robotics and digital twins as central disruptive technologies. Applications involving digital twins and simulations are central components of the industrial metaverse [18]. Changes taking place in the real world reflect the digital twin, and vice versa. Mirroring and simulating real machines, factories, cities, and other complex systems in the digital world will enable industries to solve complex real-world problems digitally and perform predictive maintenance [19].

Digitisation can be divided into three categories: digital model, digital shadow, and digital twin [20]. The digital model can be done of a physical object (e.g., a CAD model of a machine), but there is to be no automatic data exchange between the physical model and the digital model. A digital shadow is a digital representation of an object that has a one-way flow between the physical and digital object; once the physical entity changes, the digital shadow changes respectively. However, digital twins are at the core of metaverse thinking, which is that virtual and physical reality will merge into one (or at least they begin to resemble each other as much as possible). Changes in either one is to be updated automatically in the other one. To obtain a full metaverse, the digital twin must be dynamically updated according to the changes happening in physical reality. [20]

We propose that one potential way to pave the path towards viable industrial metaverse solutions is to use AMRs for autonomous geospatial digital shadow creation and updates. In addition to digital shadow creation, AMRs can collect localised data to be augmented in the digital shadow [21]. Robots, in particular social robots, are mentioned as one of the central metaverse technology enablers [5]. However, there is a lack of empirical metaverse studies, particularly any involving industrial B2B AMRs and studying their linkages to the metaverse phenomenon.

B. *AMRs and XR solutions*

By definition, Autonomous Mobile Robots (AMRs) are "industrial robots that use a decentralised decision-making process for collision-free navigation to provide a platform for material handling, collaborative activities, and full services within a bounded area" [15]. The robots detect their surroundings with multiple sensors (e.g., laser scanners and cameras), move and navigate autonomously and avoid collisions with obstacles or humans [22]. Nowadays, AMRs very commonly contain 3D Light Detection And Ranging (LiDAR) scanners for Simultaneous Location and Mapping (SLAM). By forming a point cloud, i.e., a digital 3D reconstruction of the environment, they are able to autonomously navigate on their reassigned routes or areas, and avoid collisions [22].

LiDAR-based point clouds can be regarded as up-to-date digital shadows of factories. This kind of autonomously created digital shadow has been previously used for enhanced situational awareness and change detection in factories [23]. We regard autonomous digital shadow creation as the first phase needed for industrial metaverse creation. Consequently, we propose the autonomously created digital shadow to be augmented with multimodal data visualisation and multiuser interaction modalities in an immersive factory metaverse. For these purposes, XR solutions offer wide possibilities.

XR solutions are central technological enablers in the metaverse phenomenon [3]-[5]. Research on interconnections between XR and robotics is abundant. Earlier research has addressed the following cases for VR and AR solutions for human–robot interaction: operator support, simulation, instruction, and manipulation [24]. AR and VR solutions can be used, for instance, for robot task planning [25] and mobile robot fleet visualisation [26]. Digital twins and the metaverse can serve as a virtual testing ground for new robot designs [5]. However, empirical insights on autonomous creation of factory metaverses with immersive, multiuser interaction features remain limited.

C. *Literature synthesis*

The industrial metaverse is an emerging research topic within the Production and Operations Management (POMS) literature [12][13]. However, empirical research on concrete industrial metaverse applications remains limited. More empirical insights are needed on the interconnection of AMRs and the ways they could enhance metaverse applications. In particular, more knowledge is needed on the implications of an industrial metaverse on industrial operations.

As the basis for creating a "factory metaverse", a geospatial digital shadow is needed. XR solutions, as such, do not offer a possibility for automatic and up-to-date digital shadow creation. This is something that AMRs can do. AMRs can create up-to-date digital shadows automatically with their LiDARs and cameras [23]. That way, there is always an up-to-date digital shadow available. In addition to the digital shadow creation, AMRs can collect data in the factory.

The role of AMRs has already been noted in the literature concerning central industrial metaverse technologies [27]. However, empirical studies addressing interconnections between AMRs and metaverse, and potential new industrial metaverse concepts and applications, remain scarce. Our study combines AMRs, digital shadows, data visualisation, and multi-user interaction in an immersive XR environment in the automotive industry context. This study provides empirical insights on the ways autonomous metaverse solutions can enhance industrial operations in the future.

III. METHODOLOGY

The main research question of this study is: *"How can a new AMR-assisted metaverse solution support industrial operations in the manufacturing sector?"* In order to increase understanding of a new, underinvestigated phenomenon, and to answer "how" questions, this study applies a qualitative case study approach [28]. This study combined technical development, real-life testing and demonstration, and a qualitative single case study with group interviews taking place after testing and demos. Developing and testing a new robot metaverse solution concretised the emerging phenomenon, gave concrete testing results for further technical development, and facilitated innovation of new concepts with the industrial representatives.

The study was conducted in cooperation between the VTT Technical Research Centre of Finland and Telia Finland Ltd. Telia is a large telecom company situated in the Nordic countries in Europe. Telia's main business relies on mobile networks (e.g., 5G) and telecommunications services. Their interest in this study arose from their industrial services business line and innovation of new potential future robotics and metaverse-related services. The industrial metaverse can be considered a new communication platform, therefore a core service offering for a telecom operator. Widescale deployments will also impact the industrial customers' ICT architectures as more computational capacity will be needed close to the end users and supported with low latency, high bandwidth connectivity.

In this study, VTT was responsible for the technical development work. Solution testing, demonstrations, interviews, data analysis, conclusions and reporting were conducted in cooperation between VTT and Telia. The VTT-Telia collaboration took place as a part of a large research project MURO – "Multi-purpose service robotics as an operator business" (2021-2023). Finalizing the paper and publication took place in VTT's MixedFleet project.

The AMR-enabled metaverse solution developed and studied bases on an autonomous SPOT mobile robot platform (Figure 1), the Trimble X7 LiDAR scanner and Unity game engine.

Figure 1. Autonomous scanning of Valmet Automotive's Innovation Center.

Boston Dynamics' SPOT robot represents a mobile robot with outstanding balance and mobility skills. Compared to wheeled AMRs, SPOT can move on, e.g., stairs, doorsteps, and uneven terrains. Thus, it opens up new application opportunities in many industries. In this project, the main task of the robot was to autonomously create a point cloud – in effect, a 3D geospatial digital shadow – of Valmet Automotive's Innovation Center.

In this project, the Unity platform was used for virtual world creation. Unity is a cross-platform game engine developed by Unity Technologies. It is widely used for video games, but also for professional applications in various industries. In addition to the SPOT robot and Trimble X7 scanner, a 360 camera was used. It enabled video creation in selected Points of Interest (PoIs) in the 3D model. Surrounding audio was recorded with a Zoom H4n audio recorder.

Valmet Automotive's Innovation Center was selected as the testing and demonstration site. Headquartered in Finland, Valmet Automotive is a versatile service provider for the automotive industry, one of the largest vehicle contract manufacturers in the world, and currently makes vehicles for Mercedes-Benz and Mercedes-AMG. There is a welding robot demonstration cell in the Innovation Center. It was scanned both from outside and inside in order to enable people "going inside" the robot cell in the created factory metaverse. In normal situations, naturally people are not allowed to go inside the robot cell while the robots are moving, but in the virtual world or "metaverse" people can do so virtually and remotely.

Technical development focused on building a pipeline from SPOT robot-enabled autonomous scanning to factory metaverse creation and use in VR. Technical development work proceeded in the following steps:

1. Physical and software integration between the SPOT robot platform and the Trimble scanner
2. Activation of autonomous scanning features of the SPOT&Trimble combination
3. Autonomous point cloud capture and data collection in the Valmet Automotive Innovation Center:
    - Autonomous point cloud capture with SPOT robot Autowalk functionality and Trimble X7 scanner
    - 360 video shoots
    - Audio sample recording
4. Unified point cloud creation with a Trimble FieldLink application
    - Individual scan registration to a unified point cloud
    - Addition of colour information
    - Export of the complete point cloud as an E57 file from the FieldLink
5. Point cloud processing
    - Removing of unnecessary areas and downsampling (3 mm minimum point distance) in CloudCompare
6. Development of point cloud to Unity import pipeline using the Pixyz Plugin for Unity
7. Importing the point cloud to Unity including segmentation of the point cloud, generation of Level of Detail (LOD) groups for PoIs, and a collider mesh for interaction
8. Development of the OpenXR-based device agnostic XR application platform in Unity
9. Development of user interfaces supporting XR devices, traditional keyboard and mouse interaction using the Unity XR Interaction Toolkit together with the XR Device Simulator
10. Development of multi-player features with Unity Netcode for GameObjects Software Development Kit (SDK) and Dissonance Voice Chat plugin
11. Adding 360 still images and videos to provide a detailed view of the key PoIs
12. Adding audio data in Unity with location so that the factory audio (e.g., a machine voice) always comes from the right direction for the user.

In addition to technical R&D, a qualitative case study was conducted. The purpose of the case study was to gain deeper understanding regarding the metaverse phenomenon and its potential applications in the practical industrial world. The case study included a literature review, metaverse solution demonstration in Valmet Automotive, data collection with group discussions (n=36) and interviews with a part of the group discussion and demo participants (12 informants), analysing and summarising the interview data, and drawing conclusions. In the discussions and interviews, the informants were asked to describe their initial impressions after testing the solution, and to bring up ideas for potential use cases in factory operations.

Thus, the results of this study are twofold: 1) the developed and tested new AMR-enabled factory metaverse solution and 2) the identified factory metaverse use cases and concepts.

IV. RESULTS

This Section presents the results of the study. It describes the created AMR-enabled factory metaverse solution, its features and the identified potential use cases and concepts in the automotive industry.

A. AMR-enabled factory metaverse solution

As the main result of the study, an AMR-enabled factory metaverse solution was created. The SPOT robot operated as the autonomous mobile platform for the developed metaverse-pipeline. Together with a high-speed Trimble X7 laser scanner mounted on the SPOT, autonomous scanning was made possible for generating point clouds. The rest of the development work was done in Unity. Figure 2 illustrates the pipeline from autonomous scanning, through the mesh creation process, to the virtual model:

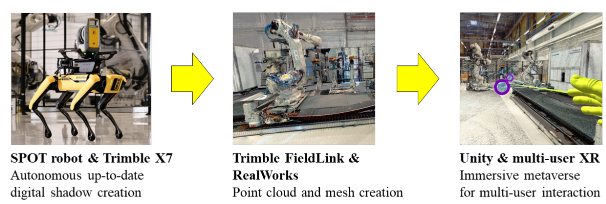


Figure 2. Pipeline from autonomous scanning to immersive, multi-user metaverse creation and utilisation.

The factory metaverse solution includes the following features:

- Immersive 3D view, factory audio and movement in an accurate and up-to-date factory metaverse
- Annotations: the possibility to add, modify and delete annotations, e.g., notes
- Versatile data visualisation and PoIs with still images and video feed
- Multi-user use, hosting and joining of a team session
- Visible avatars (head, hands and name) and interaction with a multiplayer voice chat feature
- Many user interface possibilities: can be used with XR devices, a basic keyboard or a mouse
- As XR device-agnostic as possible

 ISBN: 978-1-68558-140-4

- Annotation information in a remote database, making the annotations persistent between multiple sessions.

Users can view the factory metaverse with VR glasses or on a desktop. By clicking PoIs, the user sees still images and video feed of the robots operating. Figure 3 presents a 360 camera view and a point cloud, a digital "shadow", of the robot cell in Valmet Automotive's Innovation Center.

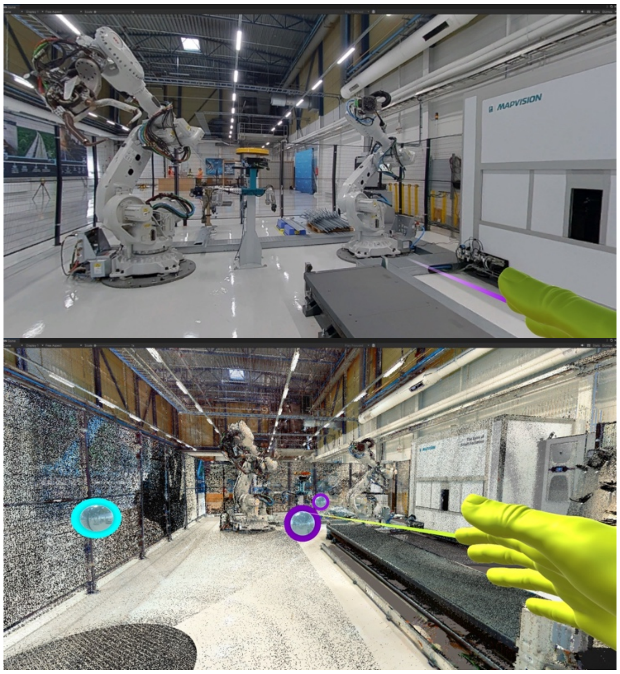


Figure 3. 360 camera view and a point cloud of the robot cell.

B. *Factory metaverse use cases and concepts*

As a result of the qualitative case study, several new AMR-enabled factory metaverse concept ideas were collected from the company representatives. The car manufacturer has a virtual R&D team distributed in several countries. They could use the autonomously created metaverse in their product design. They could represent new production techniques to their business customers in a 3D metaverse. The solution enables users to go virtually "inside" the welding robot cell while the robot is working, which would not be possible in real life for safety reasons. Then, an immersive experience would be achieved for the customers in product and manufacturing technique representations. As a director of the company pointed out regarding the related benefits: *"World-class car manufacturing competence is what we are selling. If we can concretise and visualise our competences in a better way for our clients, a superior customer experience can be offered. We will also be a forerunner in that sense."*

According to the interviewees, the solution could be used in manufacturing cell layout and process design. A virtual team comprising several organisational units, could combine their competences, plan and visualise optimal ways to conduct assembly work together within an autonomously created virtual world. As car manufacturing is a highly demanding and efficient process, all kinds of pre-planning is extremely important, when any changes are proposed for the manufacturing line. The interviewees saw that the metaverse solution brings benefits for their central need for careful design, testing and planning in everything that takes place in the factory. Also in internal logistics operations, an up-to-date 3D model could be used for material handling processes planning and optimisation.

When a factory metaverse would be created and updated autonomously with mobile robots, the 3D model could also entail information on tidiness and safety issues. The robot could autonomously and automatically detect safety risks and deficiencies and update the data in the factory metaverse. Then, the factory metaverse could also be used in safety and assembly work trainings. As in planning and R&D, trainings could also be conducted remotely and still in a very realistic way. In repair and maintenance, an autonomously created metaverse could be used for remote technical support. As an interviewee described: *"The expert wouldn't need to fly from the other end of the world, but he/she could provide technical support and advice remotely inside the shared metaverse"* What they also emphasised was the fact that all remote work and decreasing travelling supports the company's sustainability goals.

The ICT Director of the company brought up the need for an iterative process between the digital model and real world to ensure up-to-date metaverse and data. This need is at the very core of metaverse thinking according to which the real world and metaverse would closely resemble each other and updates to either one would also take place in the other "reality".

As a summary, the interviewees saw a lot of potential for the new solution and identified several operations the new solution could support. Further development needs they brought up related to the price of autonomous robots, lack of real-time updating of the model, and usability issues and simulator sickness while using VR glasses. Table 1 summarizes the identified new robot metaverse concepts and operations they support in manufacturing.

TABLE I. FACTORY METAVERSE USE CASES AND CONCEPTS

Industrial operation	**Factory metaverse use cases and concepts**
Product design	• Remote R&D team collaboration • Product design and updates visualisation
Manufacturing	• Robot and manufacturing cell design • Assembly work planning and training • New personnel onboarding and training
Internal logistics	• Planning, visualisation and optimisation of material handling
Repair and maintenance	• Machine data collection and visual display • Remote technical support from global experts
Sales and marketing	• Presenting new car designs for the customer • Demonstration and visualisation of world-class manufacturing competences for the customer
Safety	• Safety risk identification and collection autonomously and visual data display • Safety trainings

 ISBN: 978-1-68558-140-4

V. CONCLUSIONS

This Section outlines the study contribution, its limitations and potential future research avenues.

A. Study contribution and discussion

Despite the increasing interest and future business potential related to the industrial metaverse [2]-[6] and the salient role of both XR and AMR technologies as a part of the phenomenon [3][5], empirical research on new industrial metaverse solutions remains limited in the operations management domain. This study created a *pipeline from autonomous SPOT robot-enabled digital shadows to immersive, interactive, multi-user factory metaverse*, and demonstrated the solution in the automotive industry case. The study contributes by presenting the new technological solution and managerial insights into the potential use cases and future factory metaverse concepts.

XR solutions have been widely applied and studied before [3][5]. However, the novelty value of this study is derived from building the metaverse based on AMRs with LiDAR scanners and covering the whole pipeline from autonomous point cloud capture to factory metaverse creation and use. Firstly, autonomous scanning accrues several benefits. AMRs can go around a large factory and factory area and repeteadly scan the environment and collect beneficial data and identify anomalies. No manual, time-consuming scannings are then needed, and even more importantly, the digital shadow of the factory always remains up-to-date with beneficial data. Quite often, a problem with digital geospatial shadows or any sort of factory 3D models is that after they are created once, some changes take place in the factory imminently thereafter and the model becomes obsolete and useless. AMRs' autonomous regular scanning rounds tackle this problem.

Another novelty value and benefit of the solution is the fact that its basis is LiDAR-based point clouds instead of 2D pictures. The metaverse is then accurate in terms of dimensions, 3D view and experience. When something is annotated in the model, it is in the right coordination and location. You can move and interact inside the metaverse and see, e.g., a welding robot in 3D and look around it from all directions and angles. This facilitates solving technical problems together remotely in the metaverse as you can, e.g., point out a certain component specifically for which the solving of a given problem concerns the robot. You can "go inside" a welding robot cell while it is operating, which is naturally not possible in the real world for safety reasons.

The factory metaverse solution was demonstrated and tested in the Valmet Automotive Innovation Center. Company representatives identified the following use cases and concepts as having the most potential in the future: remote R&D team collaboration in product design, planning of manufacturing cells and work tasks, remote technical support work, and new personnel onboarding and safety trainings. The multi-user feature of the factory metaverse solution provides several benefits and future application opportunities by enhancing remote collaboration, ideation, and technical support, as well as knowledge sharing, visualisation and competence building over the organisational silos.

The company representatives found the solution as a promising and illustrating example of future industrial metaverse solutions. However, more clarification is needed, e.g., on the way the new solution complements and offers added value to the existing solutions already in use in the factory. Although further R&D and testing is needed, this study is one of the first empirical industrial metaverse studies with a new technical solution and demonstration complementing earlier metaverse research within the operations management domain [12][13].

B. Limitations and future research avenues

As with all research endeavours, this one had its limitations, which on the other hand also provide potential future research avenues. The first limitation concerns the fact that this was one of first attempts to build a pipeline from the SPOT robot point cloud capture to XR. All parts of the solution were developed and demonstrated successfully. However, there are still plenty of manual work phases and the pipeline does not transition automatically from scanning to a ready factory metaverse. Attempts to automate and make the solution as real-time as possible are potential avenues for future R&D.

Another limitation is that this study only included one case. Multiple case studies would provide more versatile insights both in terms of further R&D needs and potential use cases. Different use environments may hinder the use of AMRs. For instance, using the SPOT robot autonomously in public areas may entail safety risks, so it is one limitation of this specific solution. In addition, reflective surfaces and monotonous, repetitive environments may cause challenges for LiDAR scanning. Consequently, more research should be conducted in order to evaluate the viability and feasibility of the solution in different contexts and use environments.

Manifestation of an autonomous factory metaverse is a research topic that may quite certainly attract more attention and interest both within academia and the business world in the future. An AMR can be equipped with multiple sensors enhancing extensive multimodal data collection and transfer. Sophisticated computer vision adds capabilities for robots enabling identification of, e.g., anomalies and safety risks. Both robots' own status data and data they collect from the factory can be brought into the factory metaverse for enhanced situational awareness and safety. As an industrial metaverse includes many similar aspects to Industry 4.0 and 5.0, more research is needed on the overlaps, differences and linkages between the phenomena. For instance, immersive experience and multi-user collaboration are highlighted in the metaverse phenomenon. Especially from this perspective, more research is needed on the related industrial applications and potential added value for operations management.

As this study concentrated on the autonomous creation of a factory metaverse representing a digital "shadow" of the factory, more research could continue in studying feedback loops from the digital shadow back to the reality based on human and team actions taken in the metaverse. Then, we

 ISBN: 978-1-68558-140-4

will get closer to the core idea of the metaverse: the digital and real world merge and become more and more as one.

ACKNOWLEDGMENT

The MURO and MixedFleet research projects were funded by Business Finland, VTT and the participating companies.

REFERENCES

[1] N. Stephenson, *Snow crash*. New York: Bantam Books, 1992, ISBN 10: 0553351923.

[2] G. D. Ritterbusch and M. R. Teichmann, "Defining the metaverse: A systematic literature review," *IEEE Access,* 2023, https://doi.org/10.1109/ACCESS.2023.3241809.

[3] Y. K. Dwivedi *et al.*, "Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management,* vol. 66, p. 102542, 2022/10/01/ 2022, doi: https://doi.org/10.1016/j.ijinfomgt.2022.102542.

[4] S. M. Park and Y. G. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE Access,* vol. 10, pp. 4209-4251, 2022, doi: 10.1109/ACCESS.2021.3140175.

[5] L.-H. Lee *et al.*, "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," *arXiv preprint arXiv:2110.05352,* 2021.

[6] E. Dincelli and A. Yayla, "Immersive virtual reality in the age of the metaverse: A hybrid-narrative review based on the technology affordance perspective," *The Journal of Strategic Information Systems,* vol. 31, no. 2, p. 101717, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.jsis.2022.101717.

[7] G. Lawton, "Why the industrial metaverse will eclipse the consumer one," https://venturebeat.com/virtual/why-the-industrial-metaversewill-eclipse-the-consumer-one/ (accessed: Jan. 29, 2024).

[8] MIT Technology Review, "The industrial metaverse: a game-changer for operational technology," https://www.technologyreview.com/2022/12/05/1063828/the-industrial-metaverse-a-game-changer-for-operational-technology/ (accessed: Jan. 29 2024).

[9] E. Hazan, G. Kelly, H. Khan, D. Spillecke, and L. Yee, "Marketing in the metaverse: An opportunity for innovation and experimentation," McKinsey & Company. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/marketing-in-the-metaverse-an-opportunity-for-innovation-and-experimentation (accessed: Jan. 29, 2024).

[10] M. Purdy, "How the metaverse could change work," Harvard Business Review. https://hbr.org/2022/04/how-the-metaverse-could-change-work (accessed: Jan. 29, 2024).

[11] K. B. Ooi *et al.*, "The metaverse in engineering management: Overview, opportunities, challenges, and future research agenda," *IEEE Transactions on Engineering Management,* pp. 1-8, 2023, doi: 10.1109/TEM.2023.3307562.

[12] M. M. Queiroz, S. Fosso Wamba, S. C. F. Pereira, and C. J. Chiappetta Jabbour, "The metaverse as a breakthrough for operations and supply chain management: implications and call for action," *International Journal of Operations & Production Management,* vol. ahead-of-print, no. ahead-of-print, 2023, doi: 10.1108/IJOPM-01-2023-0006.

[13] Y. Ukhanov and A. Berggren, "Exploring the potential of the metaverse in operations management: Towards sustainable practices," ed, 2023, ISRN: JU-JTH-PRS-2-20230126.

[14] K. Salminen, M. Kivinen, and A. Kokkonen, "Handbook for human-driven industrial metaverse," VTT Technical Research Centre of Finland. https://info.vttresearch.com/download-handbook-of-human-driven-industrial-metaverse (accessed: Jan. 29, 2024).

[15] G. Fragapane, R. de Koster, F. Sgarbossa, and J. O. Strandhagen, "Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda," *European Journal of Operational Research,* vol. 294, no. 2, pp. 405-426, 2021/10/16/ 2021, doi: https://doi.org/10.1016/j.ejor.2021.01.019.

[16] W. Knight, "BMW's virtual factory uses AI to hone the assembly line," Wired. https://www.wired.com/story/bmw-virtual-factory-ai-hone-assembly-line/?utm_source=EU+Automation&utm_medium=PR&utm_campaign=EUA1538 (accessed: Jan. 29, 2024).

[17] A. Kusiak, "Manufacturing metaverse," *Journal of intelligent manufacturing,* pp. 1-2, 2023, doi: https://doi.org/10.1007/s10845-023-02145-w.

[18] B. Starly, P. Koprov, A. Bharadwaj, T. Batchelder, and B. Breitenbach, "“Unreal” factories: Next generation of digital twins of machines and factories in the Industrial Metaverse," *Manufacturing Letters,* vol. 37, pp. 50-52, 2023/09/01/ 2023, doi: https://doi.org/10.1016/j.mfglet.2023.07.021.

[19] "The emergent industrial metaverse," https://www.technologyreview.com/2023/03/29/1070355/the-emergent-industrial-metaverse/ (accessed: Jan. 29, 2024).

[20] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE access,* vol. 8, pp. 108952-108971, 2020, doi: http://dx.doi.org/10.1109/ACCESS.2020.2998358.

[21] T. Hakanen and M. Lepola, "Taking multi-purpose 5G robots to industrial metaverse," VTT Technical Research Centre of Finland. https://www.murorobotics.fi/post/taking-multi-purpose-5g-robots-to-industrial-metaverse (accessed: Jan. 29, 2024).

[22] D. De Silva, J. Roche, and A. Kondoz, "Robust fusion of LiDAR and wide-angle camera data for autonomous mobile robots," *Sensors,* vol. 18, p. 2730, 08/20 2018, doi: 10.3390/s18082730.

[23] T. Hakanen, P. Kemppi, and P. Tikka, "Mobile 3D LiDAR-based object and change detection in production and operations management," *ICAS 2023,* p. 10, 2023, ISBN: 978-1-68558-053-7.

[24] M. Dianatfar, J. Latokartano, and M. Lanz, "Review on existing VR/AR solutions in human–robot collaboration," *Procedia CIRP,* vol. 97, pp. 407-411, 2021/01/01/ 2021, doi: https://doi.org/10.1016/j.procir.2020.05.259.

[25] H. Fang, S. Ong, and A. Nee, "A novel augmented reality-based interface for robot path planning," *International Journal on Interactive Design and Manufacturing (IJIDeM),* vol. 8, pp. 33-42, 2014, doi: https://doi.org/10.1007/s12008-013-0191-2.

[26] J. Roldán Gómez *et al.*, "Multi-robot systems, virtual reality and ROS: Developing a new generation of operator interfaces," 2019, pp. 29-64, doi: https://doi.org/10.1007/978-3-319-91590-6_2.

[27] N. Kshetri, "The economics of the industrial metaverse," *IT Professional,* vol. 25, no. 01, pp. 84-88, 2023, doi: 10.1109/MITP.2023.3236494.

[28] R. K. Yin, *Case study research design and methods (5th ed.)* (The Canadian Journal of Program Evaluation). 2014, p. 282, doi: http://dx.doi.org/10.3138/cjpe.30.1.108.

DebiAI: Open-Source Toolkit for Data Analysis, Visualisation and Evaluation in Machine Learning

Tom Mansion
IRT SystemX
91120 Palaiseau, France
tom.mansion@irt-systemx.fr

Raphaël Braud
IRT SystemX
91120 Palaiseau, France
raphael.braud@irt-systemx.fr

Ahmed Amrani
IRT SystemX
91120 Palaiseau, France
ahmed.amrani@irt-systemx.fr

Sabrina Chaouche
IRT SystemX
91120 Palaiseau, France
sabrina.chaouche@irt-systemx.fr

Faouzi Adjed
IRT SystemX
91120 Palaiseau, France
faouzi.adjed@irt-systemx.fr

Loïc Cantat
IRT SystemX
91120 Palaiseau, France
loic.cantat@irt-systemx.fr

***Abstract*—DebiAI is an open-source tool designed for data analysis, visualization, as well as evaluation and comparison of Machine Learning (ML) models. It is intended to be used both at the stage of the project data preparation, and for the evaluation of the ML models performances. It has a rich and user-friendly graphical interface that allows to visualize, analyze, select, edit and annotate data and metadata, as well as for bias detection and contextual evaluation of ML models. The tool relies on a generic data model, making it applicable to any type of ML task: classification, regression, object detection in images and more. It is an open source code distributed under the Apache License, Version 2.0 . The code is publicly available on https://github.com/debiai and further information along with guidelines for the users can be found on its dedicated website https://debiai.irt-systemx.fr.**

Index Terms*—*Data Analysis; Data Visualization; Bias Detection; Contextual Evaluation; Machine Learning; Trustworthy AI.

I. Introduction

Data analysis and visualization are fundamental tasks in Machine Learning (ML) projects. They are playing a crucial role in a typical ML process, and are contributing not only in the data preparation phase but also during and after the model building. This emerging research topic, which combines several interactive systems and domains for ML processes, focused on human interaction and collaboration [1] created a new field that is Human-Centred Machine Learning (HCML) interaction [2]. Thus, a typical HCML framework allows an interactive visual analytic and ML evaluation [3]. Consequently, efficient tools are needed to assist, during the whole ML process, in data preparation and quality inspection before and after using the data for model training.

An effective tool should help enhance the iterative ML process across various phases: data preparation, analysis, anomaly detection, and annotation during the preparation phase, as well as the analysis and evaluation of models results. This aids in identifying the model's weaknesses and detecting issues at the data level. In real-world ML projects (e.g., industrial ones), the data is enriched by metadata including operational context, expert knowledge, etc; that can help better understand the raw data and can be informative for the learning process, and hence improving the quality of the model's learning and predictions. Having such tools contribute also to the enhancing the trustworthiness of the used ML algorithm.

In ML-based engineering systems, it is crucial to guarantee key properties like accuracy, robustness, explainability, fairness, privacy, among many other primary values of Trustworthiness. Current research and development challenges of deploying trustworthy ML solutions are covered by wide programs like Confiance.ai [4], the French AI flagship program to industrialize trustworthy AI-based critical systems [5] [6] and the TAILOR [7] network at the European level.

DebiAI has been developed by the IRT SystemX in the framework of Confiance.ai program to contribute in ensuring trustworthiness by data, and serves as the main interface to view, analyze, select, edit and/or annotate any type of data and metadata. The rest of the paper is organized as following: Section II presents a brief state of the art of developed HCML tools, Section III presents methodology, followed by implementation description and application in a real use case described in Sections IV and V, respectively. Finally, conclusions and perspectives are drawn in Section VI.

II. Literature review

Data visualization is the practice of representing information using graphical representations, employing technology-driven tools and software. Its fundamental objective is to enhance pattern recognition, improve understanding of complex concepts and facilitate in-depth exploration, thereby fostering the generation of new insights. Well-designed data visualizations can help in understanding large datasets and establishing connections between ideas, concepts, and processing stages. Therefore, visual analysis can contribute to the optimization of AI approaches by actively participating in all aspects of the model building process [8] [9]. Similarly, Hohman et al. [10] highlights that successful ML applications often require iterations in data handling and continuous adjustments of the

 ISBN: 978-1-68558-140-4

model. The authors introduced CHAMELEON, an interactive tool designed to attribute data iteration, thereby enhancing model performance, data validation, and the overall quality of ML projects. To improve data quality, Kandel et al. [11] presented Profiler, a tool using data mining to automatically detect issues and recommend coordinated visualizations for context-based assessment. Profiler offers methods for integrated statistical and visual analysis and view suggestions. ScrutinAI [12] is a Visual Analytics tool specifically tailored for enhancing the comprehension of deep neural network (DNN) predictions. Its primary objective is to identify and investigate potential weaknesses within models. To facilitate this, ScrutinAI provides interactive visualizations of input and output data, along with interactive plots and data filtering for comprehensive analysis of predictions. This tool is specifically designed for object detection and semantic segmentation, whereas DebiAI is applicable to a wide range of use cases. Zhang et al. [13] presented Manifold, a visual analytics platform designed for comparing and debugging ML models. The platform empowers users to categorize instances based on the model's accuracy and confidence, identify symptomatic instances that generate incorrect results and continually help to enhance the model's performance. As DebiAI, Manifold is created as a generic tool that operates independently of the internal logic of the ML model. It focuses on the input and the output. In order to improve the models' performances and gain insights into their limitations, relying solely on the overall result from the test and training sets is insufficient. To address this limitation, the ModelTracker [14] tool offers instance-level visualization of results and enables users to individually inspect each instance. The tool has been applied to the binary classification task. In [15], the authors applied the approach for a multi-class classification task. They used Parallel Coordinates Plots (PCP) to visually represent the multi-class predictions for a subset of instances. To complete these shortcomings and improve the understanding of models results, the proposed DebiAI enables the analysis of model's outcomes across various levels of granularity (instance, subset and dataset). This functionality has been applied to multiple tasks such as regression, classification, object detection, etc. In line with the approach outlined in [15], DebiAI leverages the use of PCP for model's results analysis. However, the implementation of PCP within DebiAI is flexible, enabling its use not only for result analysis but also for assessing attributes.

III. Methodology

DebiAI is a web-based visual analytics application designed to support ML and data analysis. Its emphasis lies in two crucial phases of the ML pipeline: pre-model and post-model building. It facilitates the development of ML models by assisting in data analysis during data curating and processing stage and enabling models performances comparison.

In the pre-model construction phase, DebiAI serves as a key resource for data scientists and ML engineers during project preparation. It enables them to visually identify biases and errors in data inputs, detect anomalies and outliers throughout the data life cycle, assess data quality and domain coverage through relevant metrics and select and analyze subsets of data to improve the quality of ML models.

In the post-model building phase, DebiAI serves as a visual analytics solution, simplifying the interpretation of the ML model's outputs. Its primary objective is to present the model's results in an intuitive and easily understandable manner, ultimately enhancing user confidence in the model's predictions. Additionally, DebiAI offers features to identify model's weaknesses, comparing performances, and evaluating model's effectiveness. This comprehensive approach fosters ongoing model refinement, tailored to the specific needs of the use case.

In both phases, DebiAI provides tools for creating and sharing statistical visualizations of the project data and results with collaborators (team or/and clients).

A. Functional Description

DebiAI is an intuitive visualization tool designed to simplify the creation of interactive dashboards, empowering users without little to no programming skills. It offers a diverse set of graphical widgets, including charts, tables, parallel categories, parallel coordinate plots, interval plots, night stars plots, and sample arrays. Moreover, DebiAI provides a user-friendly and flexible solution for interactive dashboard design, allowing users to effortlessly configure, adjust, resize, and position these widgets within their dashboards, ensuring the utmost customization of data presentation. This includes the ability to generate and share statistical visualizations of project data with team members or clients, fostering collaboration and informed decision-making by providing clear insights into the data. One of DebiAI's standout features is its dynamic data selection and filtering capabilities, encouraging continuous exploration. Users can effortlessly create data subsets (selections) and apply filters based on various variables and contexts. This ensures that the dashboard consistently presents the identified subset of data. Furthermore, DebiAI assists users in identifying biases and inaccuracies in inputs, results, project data contexts or ground truths, thereby improving data integrity. DebiAI facilitates the evaluation and comparison of ML model's performances within the whole dataset or a specific data subset. It enables the analysis of results across multiple levels of granularity. Indeed, the model's performances are calculated at the level of each instance. Consequently, it is possible to identify the contexts or a combination of contexts in which the model encounters difficulties. It also simplifies the generation and organization of datasets, supporting in-depth analysis and potential retraining. DebiAI relies on a generic data model that facilitates seamless application across various datasets, data types and use cases while maintaining consistent data processing practices. This essential feature provides DebiAI with flexibility, allowing it to transition effortlessly between various datasets or model's results. In addition to its visualization capabilities, DebiAI incorporates implementations of statistical measures such as correlation analysis using Pearson or Spearman coefficients. To support these visualizations, DebiAI also integrates techniques for discretizing continuous variables. In addition, it enables the use of internal or external algorithms to compute metrics or indicators on the data. Consequently, these metrics can be calculated either before integrating the data into DebiAI or during the data analysis phase. Various types of calculations can potentially be carried out by these algorithms, including the computation of new features, the assessment of model's results quality, as well as indicators of data quality and distribution.

IV. Implementation Description

In this section, we describe DebiAI's global architecture and further dive into the detail of each component.

A. DebiAI Technical architecture

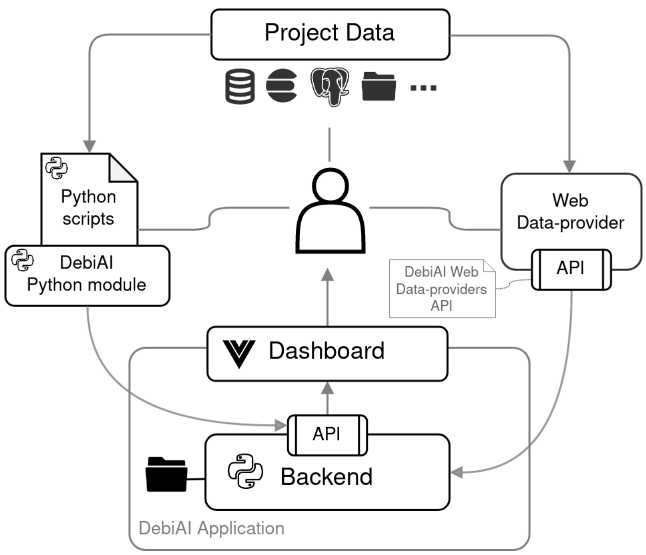


Fig. 1: DebiAI architecture overview

The architecture of DebiAI is bifurcated into two main environments (Fig. 1):

The project environment :

- Project Data: This is the source of data that the user intends to analyze. It may originate from a variety of sources and formats, such as CSV or JSON.
- Python Scripts and DebiAI Python Module: Using the DebiAI Python module, users can adapt their existing scripts and workflows to create selections and insert data and model's results into DebiAI.
- Web Data-Provider: These are the services created by the user's project that enable DebiAI to fetch data directly from the project's data sources. A Web Data-Provider can be developed using any programming language, access data from any type of database, and be hosted on any server. The only stipulation is that it should implement and expose a specific REST API according to a defined contract.

DebiAI application environment:

- DebiAI Web Dashboard: This is the user interface of DebiAI, developed using VueJs. It provides users with an interactive platform to manage and view their data, and is hosted and served by the DebiAI backend.
- Backend and API: This is a Python-powered backend that not only provides an API but also serves the Web dashboard. This API is employed by the dashboard for data retrieval and by the Python module for data insertion. Additionally, it manages communications with the Web Data-providers and processes computational requests made by the dashboard.
- Data storage: DebiAI uses a folder-based data store that contains data in a JSON format. This data store supports the DebiAI backend by retaining projects created by the Python module and some specific dashboard elements, including layout configurations for project dashboards.

B. DebiAI Generic Data Model

The data format required must follow a CSV structure with no missing values. One of the most important features of DebiAI is its data model. The main objective is to enable the determination of the format of instances and the relationship between instances, models, models' outputs, and models' evaluation metrics per instance. Syntactically, each instance is composed of attributes, contexts and annotations. The instance is linked to multiple ML models, where each model produces an output. Evaluation metrics are also associated with the model's outputs. This structure is applicable to all types of data and ML tasks (classification, regression, object detection, etc.).

C. DebiAI Data Integration Process

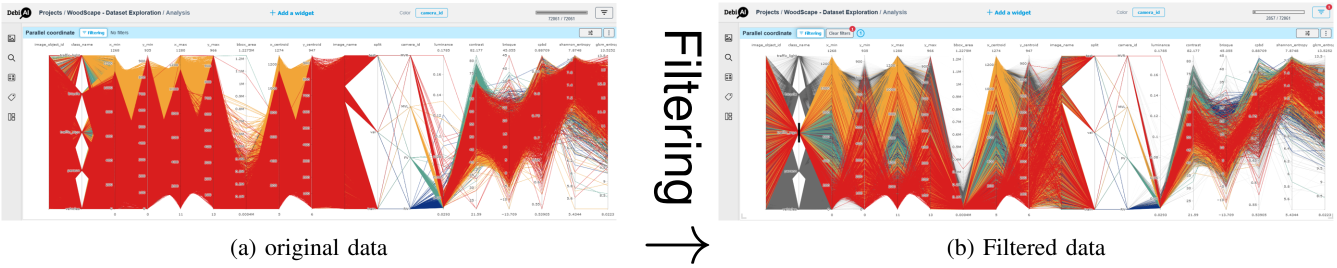
DebiAI offers two main ways to add data, each suited for different types of users and projects:

1) **Python Module:** This principal method enables seamless integration of project's data into DebiAI via a dedicated Python module. Made for an integration within Python workflows, this approach, for example, facilitates the direct transfer of models' results post-evaluation. This method is especially handy for those who primarily use Python.
2) **Data Providers:** Alternatively, DebiAI can interface with data through RESTful services, termed 'Data Providers'. This method is database-agnostic, allowing DebiAI to directly request project's data, thereby making the data loading process faster and more efficient. Unlike the Python module, it doesn't require DebiAI to duplicate data within its integrated database. Although setting up a Data Provider is more time-consuming than using the Python module, it offers greater efficiency and flexibility. This is particularly beneficial for long-term projects that regularly update their data.

Each method offers distinct benefits, and the choice depends on the specific requirements and scale of the project.

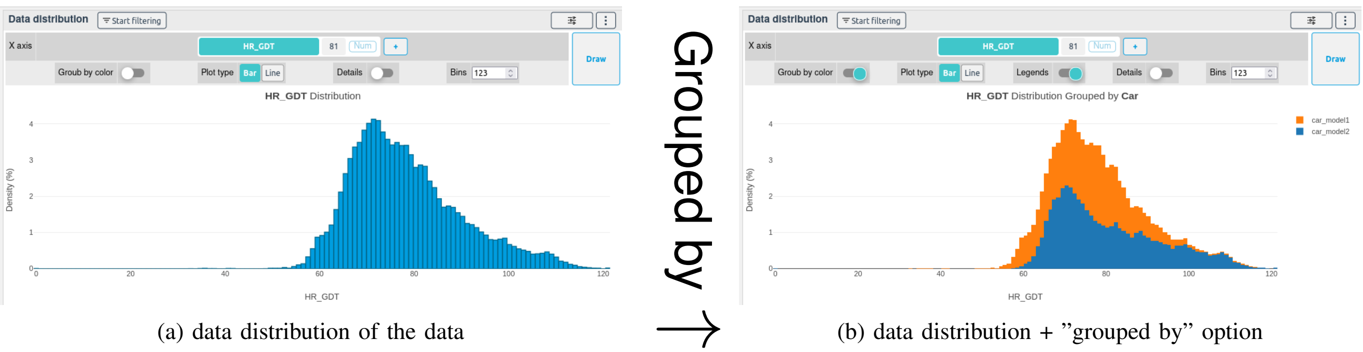
V. DebiAI Application

DebiAI is built upon a generic data model, and does not depend on data type (for instance images and time series),



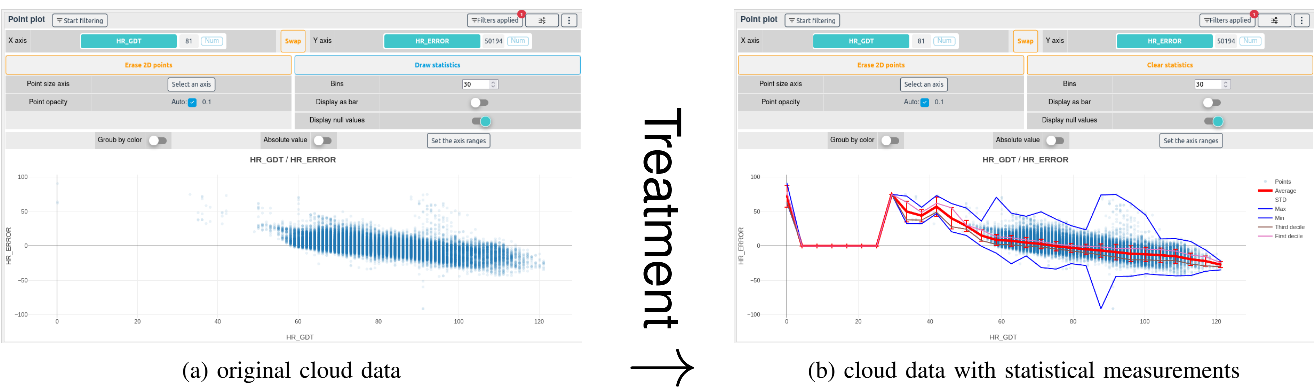
(a) original data (b) Filtered data

Fig. 2: parallel coordinates widget. (a) represents the original data uploaded and (b) represents the same widget by selecting a subset of variables interactively.



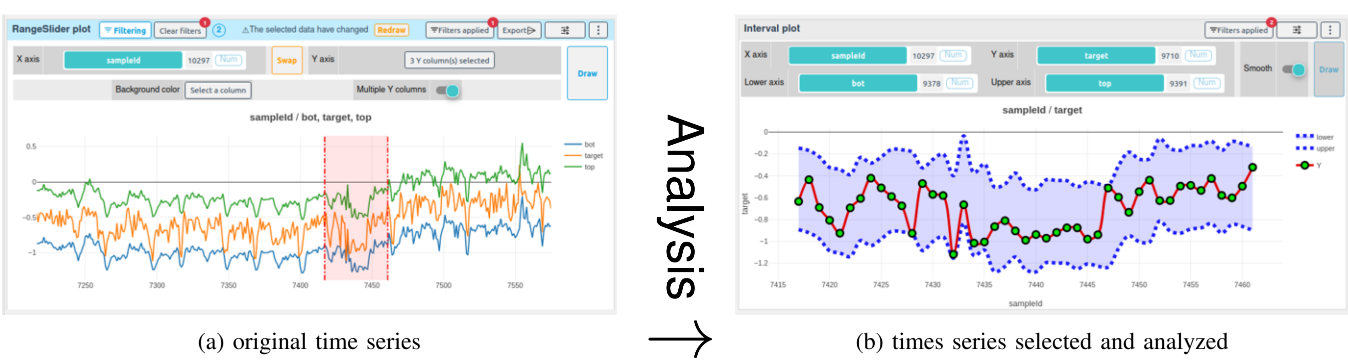
(a) data distribution of the data (b) data distribution + ”grouped by” option

Fig. 3: illustration of data distribution by adding the option of ”grouped by”. (a) represents an example of data and (b) represents the same data grouped by another variable with two different colors.



(a) original cloud data (b) cloud data with statistical measurements

Fig. 4: Statistical treatment for an example of cloud data. (a) represents the original data and (b) represents the data by adding a set of statistical measures. Here illustrated measures are: mean, standard deviation, min and max and deciles



(a) original time series (b) times series selected and analyzed

Fig. 5: Statistical analysis applied on times series example. (a) represents the original time series (b) represents the times series filtered and analyzed by adding a set of statistical measures

making it pertinent to various use cases across a multitude of datasets. This intrinsic adaptability allows it to be valuable in a wide range of scenarios. It demonstrates its utility in the analysis of time series data, simplifying essential tasks such as regression. Furthermore, its functionality seamlessly extends to computer vision applications. Indeed, DebiAI provides tailored visual support for each stage of the process, enhancing models in tasks such as object detection and image classification.

In the following two sections, we will present the use of multiple widgets in DebiAI for various use cases and provide an overview of a use case related to 2D object detection.

A. *DebiAI Visual Functionalities*

As described in Section III-A, DebiAI gives the ability to visualize and create interactive dashboards. Moreover, it can visualize various data types such as time series, point clouds and tabular data and display computed attributes of images. However, for images viewing, it can establish links with external tools. In this section, we review a set of graphics implemented on different datasets with different data types. We also illustrate the main filtering features proposed by DebiAI. Four graphical visualizations are presented by exploring the parallel coordinates, the data distribution, the points plots and the time series widgets enhanced with interactive options. Fig. 2 illustrates a visualization of a dataset by using a parallel coordinates and the possibility to filter directly a set of variables. Another graphic visualization to analyse data distributions variables with the possibility of grouping by other variables is shown in Fig. 3. The third visualization selected from DebiAI is the possibility to apply statistical measures. Fig. 4, captures a data cloud visualized with its primary statistical measures; an envelope of min and max of the data, the average, the confidence interval $\pm\sigma$, where σ represents the standard deviation and also two deciles of the data. Fig.5 visually encapsulates the two distinct stages in the statistical analysis of a time series. Initially, Fig.5a displays the original data over an extended period. Subsequently, in the second stage, Fig.5b illustrates the time series after a more detailed analysis and filtering, focusing on a shorter timeframe. Among the statistical measures, a noteworthy one to explore when analyzing dataset's variables is examining their correlation, a task effortlessly accomplished using the DebiAI's correlation matrix widget.

B. *WoodScape Dataset: 2D Object Detection Case Study*

The WoodScape dataset [16] is a public dataset containing more than 100.000 images of urban scenes captured using fish-eye cameras for automotive driving tasks. The images are provided by 4 different cameras with different angles of view (front, rear, middle right and middle left) with 360° coverage and have annotations for a diverse set of computer vision tasks.

In our study, we focus on the 2D bounding boxes detection task with five classes: vehicle, person, bicycle, traffic light and traffic sign. We applied two versions of YOLO-based architectures, specifically YOLOv5 and YOLOv8. A YOLOv5 and a YOLOv8 models trained on WoodScape dataset and a YOLOv8 model pre-trained on COCO2017 [17].

The first step is to obtain a comprehensive overview of the data distribution, understand its scope and how it can be effectively used in a ML process. This comprehension is crucial, as it helps in the formulation of effective training strategies. The Fig. 6 shows the distribution of the dataset's train set composition of each class of the five cited above grouped by the camera id. By applying the same configuration to display the distribution of each of the three sets (train, validation and test), we observe a similar distribution among the three of them. This confirms the appropriate split of the data. Nevertheless, the figure highlights the imbalance among the distributions of the five classes, suggesting the need for adaptive training techniques, for example using a weight sampler. It is essential to consider this imbalance when interpreting the models outputs to avoid biased and skewed conclusions.

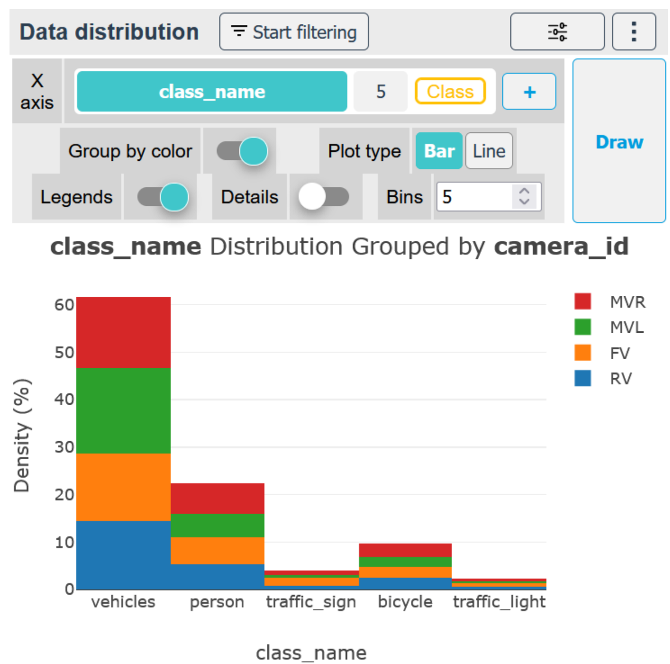


Fig. 6: Dataset's train set split by objects class distribution grouped by Camera ID

In the second stage, we used DebiAI to analyze the results of our models applied on the WoodScape test set and put them back into the context of the dataset and its features. This approach ensures an accurate interpretation of the models' outputs and provides potential improvements directions. The Fig. 7 shows the distribution of the f1-score of each model grouped by the camera id, where we can easily spot the gap in performance between the three models: having the two models trained on WoodScape dataset showing higher scores compared to the one pre-trained on COCO 2017 dataset, which is expected giving the discrepancy between the two datasets. We can also notice that the Yolov5 trained on the WoodScape

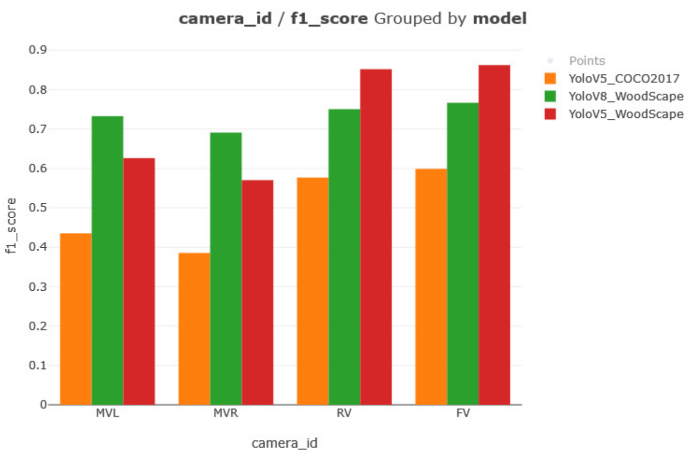


Fig. 7: f1-score results by Camera ID grouped by Models

train set has better score on the data coming from the front and rear view cameras (FV & RV) of the vehicle while the Yolov8 also trained on WoodScape shows a greater score on the middle view cameras (left & right) data. This first observations led to further investigations using DebiAI in an attempt to understand the models' outputs; you can check our demo on this case study on our website for more details.

VI. Conclusions and Perspectives

In this paper, we introduced DebiAI, a versatile web-based visual analytics tool that enhances data preparation process, quality assessment, model results analysis and comparison in ML projects. Its adaptability to various use cases and user-friendliness make it a valuable asset contributing to the trustworthiness in AI. For instance, we illustrated its application in a use case of 2D object detection task for driving assistance. As Machine Learning evolves, DebiAI can play a pivotal role in ensuring reliable and interpretable ML outcomes, solidifying its relevance in the field. In DebiAI's outlook, the priorities are to simplify the data integration process, enhance interoperability with the learning process to retrieve and analyze data from each cycle. The concepts of robustness and explainability are also tied to model's quality. Therefore, incorporating these metrics into the process is critical for overall trust. Furthermore, we are considering coupling DebiAI with a data version control system to ensure traceability of dataset and models' results evolution.

Acknowledgment

This work has been supported by the French Government under the "France 2030" program, as part of the SystemX Technological Research Institute. This work was conducted as part of the Confiance.ai program, which aims to develop innovative solutions for enhancing the reliability and trustworthiness of industrial AI-based systems.

References

[1] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim, "A Survey of Human-Centered Evaluations in Human-Centered Machine Learning," *Computer Graphics Forum*, vol. 40, no. 3, pp. 543–568, Jun. 2021, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org/10.1111/cgf.14329

[2] T. Kaluarachchi, A. Reis, and S. Nanayakkara, "A review of recent deep learning approaches in human-centered machine learning," *Sensors*, vol. 21, no. 7, p. 2514, 2021.

[3] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim, "What you see is what you can change: Human-centered machine learning by interactive visualization," *Neurocomputing*, vol. 268, pp. 164–175, 2017.

[4] Confiance.ai, "Confiance.ai Program," https:www.confiance.ai, 2024, online; accessed 19-February-2024.

[5] M. Adedjouma, C. Alix, L. Cantat, E. Jenn, J. Mattioli, B. Robert, F. Tschirhart, and J.-L. Voirin, "Engineering Dependable AI Systems," in *17th Annual System of Systems Engineering Conference (SOSE)*. Rochester, United States: IEEE, Jun. 2022. [Online]. Available: https://hal.science/hal-03700300

[6] M. Juliette, L. R. Xavier, B. Bertrand, C. Loic, T. Fabien, R. Boris, G. Rodolphe, and N. Yves, "Ai engineering to deploy reliable ai in industry," *AI4I*, vol. 25, no. 1, 2023.

[7] Taylor, "TAYLOR Network Program," https://tailor-network.eu/, 2024, online; accessed 19-February-2024.

[8] J. Yuan, C. Chen, W. Yang, M. Liu, J. Xia, and S. Liu, "A survey of visual analytics techniques for machine learning," *Computational Visual Media*, vol. 7, no. 1, pp. 3–36, Mar. 2021. [Online]. Available: https://doi.org/10.1007/s41095-020-0191-7

[9] J. Wang, S. Liu, and W. Zhang, "Visual analytics for machine learning: A data perspective survey," *CoRR*, vol. abs/2307.07712, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2307.07712

[10] F. Hohman, K. Wongsuphasawat, M. B. Kery, and K. Patel, "Understanding and visualizing data iteration in machine learning," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–13.

[11] S. Kandel, R. Parikh, A. Paepcke, J. M. Hellerstein, and J. Heer, "Profiler: Integrated statistical analysis and visualization for data quality assessment," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 547–554.

[12] E. Haedecke, M. Mock, and M. Akila, "ScrutinAI: A visual analytics tool supporting semantic assessments of object detection models," *Computers & Graphics*, vol. 114, pp. 265–275, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009784932300105X

[13] J. Zhang, Y. Wang, P. Molino, L. Li, and D. S. Ebert, "Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 364–373, 2019.

[14] S. Amershi, M. Chickering, S. Drucker, B. Lee, P. Simard, and J. Suh, "ModelTracker: Redesigning Performance Analysis Tools for Machine Learning," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2015)*. ACM - Association for Computing Machinery, Apr. 2015, edition: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2015). [Online]. Available: https://www.microsoft.com/en-us/research/publication/modeltracker-redesigning-performance-analysis-tools-for-machine-learning/

[15] D. Ren, S. Amershi, B. Lee, J. Suh, and J. D. Williams, "Squares: Supporting interactive performance analysis for multiclass classifiers," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 01, pp. 61–70, jan 2017.

[16] S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D. O'Dea, M. Uricár, S. Milz, M. Simon, K. Amende *et al.*, "Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9308–9318.

[17] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015.

A Mobile App for Exploring Chemical Molecules:

Machine Learning-Powered Handwritten Compound Identification and 3D Visualization

Luiza Coelho de Souza, Pedro Henrique Peraçoli Pereira Ceccon,
Raphael Ferezin Kitahara, Silvana Rodrigues Alves

Eldorado Institute of Research
Campinas, Brazil
email: raphael_kitahara@hotmail.com

***Abstract*—The study of Chemistry is part of the mandatory school curriculum in Brazilian basic education and is considered by students to be a difficult subject to understand and abstract, generating resistance in learning, assimilation of concepts, and applicability in everyday life. As an experimental science, laboratory practice has contributed to students' learning. However, it is not always possible to carry out experiments, as many schools do not have the necessary physical requirements or for classes that are taught online. This article presents the development of a mobile application that uses machine learning to improve the process of teaching chemistry, making it possible to identify hand-drawn molecules and display 3D virtual correspondents, showing the structure of the element as well as information that relates it to everyday life.**

Keywords - Chemistry; education; machine learning; mobile app; CoreML.

I. Introduction

The important role that technological resources have played in our contemporary society is indisputable. Nowadays, mobile devices are widely used, facilitating all areas of daily life, and significantly evolving into practically inseparable extensions of our bodies and our minds. In this context, the crucial importance of integrating them into the teaching and learning process is highlighted, especially given that, in today's day and age, young students are born and grown in an intrinsically technological society [13].

However, despite the conveniences offered by a technological world, certain academic subjects, especially chemistry, remain a great challenge to many students. In the classroom setting, chemistry education often unfolds through static representations on chalkboards or whiteboards, where intricate molecular structures are sketched in abstract notations and symbols. While these methods serve as initial blueprints, they also pose a difficulty for students, given that it is inherently challenging to understand and comprehend a complex subject solely through static drawings and abstract concepts, added to the fact that they cannot see it in their daily lives or understand how it affects them [14].

This paper shows the development process of a mobile app aimed at helping students with the challenges of learning chemistry, with the integration of technology. It benefits their learning process, while also helping them identify drawn 2D static molecules presented to them in the classroom and understand the meaning and importance of it. By bridging this gap between abstract concepts and the real-world, the application aims to make the study of chemistry more accessible and engaging, offering an interactive alternative solution to understand the subject and appreciate the presence of these structures in their everyday lives.

The remaining of the paper is structured as follows. Section 2 presents how the use technology enhances chemistry education and addresses learning challenges. Section 3 presents the topic of supervised machine learning and CNNs to classify hand-drawn chemical molecules. Section 4 presents the development of an educational app for learning chemistry through machine learning-powered molecule recognition. Section 5 presents the developed application, an interactive app for identifying and learning about hand-drawn chemical molecules. Section 6 presents the possible empowerment of Brazilian education with a machine learning app for interactive chemistry learning.

II. Chemistry Education and Visual-Based Learning

The challenges faced during the teaching and learning process often result in significant failure rates and a growing lack of interest on the part of both students, who face learning difficulties, and teachers, who feel challenged to impart knowledge [1]. This difficult challenge is particularly noticeable in the teaching of chemistry, which can be attributed to traditional methods that combined with complex content, make classes monotonous and discouraging.

The study of chemistry poses a challenge in visualizing molecule structures and their complexities, especially in settings where laboratories are scarce and access to visual aids and interactive learning is not guaranteed. In 2019, only 42,1% of public high schools in Brazil had Science laboratory structures, while, in public middle schools, that number got as low as 8,6% [7]. Furthermore, considering the Brazilian standardized high school exam (ENEM), which can be used to understand better students' performance on mandatory subjects, data shows that candidates performed below 50% in questions that include chemical equations and symbols (INEP, 2019), which alerts to the alarming situation regarding the subject and the importance of offering learning alternatives.

Various approaches have been proposed to overcome the learning difficulties, especially in the field of Chemistry.

These strategies range from using scientific articles as teaching resources to carrying out practical experiments with everyday materials, to using different technologies to improve visuals and get students' attention [2].

The Brazilian Ministry of Education launched the Education Technology Guide in 2008 which describes a series of options for teachers to use in the classroom to enrich the pedagogical process. The use of technology is of great importance to the teaching and learning process, and this is happening rapidly. In 2019, the percentage of people who used the internet was 88.1% among students and 75.8% among non-students. In addition, the study found that among the population aged 10 and over, the main means of access was the cell phone (98.6%), followed by the microcomputer (46.2%), television (31.9%) and tablet (10.9%) [6].

Efforts are also being made to develop new pedagogical approaches that improve students' understanding since a large part of the topics studied in chemistry surround the understanding of chemical elements and their interactions. Although most of the elements are essentially three-dimensional, they are often presented in two-dimensional form through graphs and projections, which are difficult to understand and do not pose as a solution to the comprehension problem faced by students today [3]. The identification of chemical compounds is fundamental in the field of chemistry. However, the process of identifying these compounds can be a difficult task, especially when handwritten.

Considering this scenario, mobile applications serve to integrate technology into the students' learning process, given that they can be accessed by most students with devices within their reach. This could mean a positive scenario when considering the difficulties with students within chemistry, given that not only does mobile integration promote students' learning but it also affects their motivation [8].

III. MACHINE LEARNING

Machine learning is a constantly evolving sector among computational algorithms designed to emulate human intelligence by learning from the surrounding environment. They are considered the workhorse of the new era of so-called big data. Machine learning-based techniques have been applied in diverse fields, from pattern recognition, computer vision, spacecraft engineering, finance, and computational biology to biomedical and medical applications [4].

The intersection of machine learning and molecular chemistry education offers significant potential to improve teaching effectiveness by making learning more engaging, accessible, and tailored to individual needs. This innovative approach promises to transform the way concepts of molecular chemistry are understood in educational settings [5]. An example of the current use of machine learning in an educational setting is the identification of two-dimensional structure of mathematical symbols and expressions [15].

Among the variety of machine learning models, Supervised learning, the task function that maps an input to an output based on example input-output pairs, is widely used in detection models. It infers a function from labeled training data consisting of training examples. The basic flow of dynamics that occurs in supervised machine learning algorithms can be seen in Figure 1.

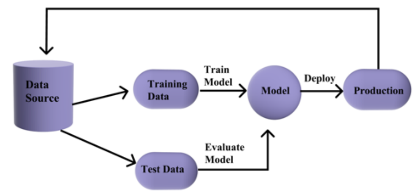


Figure 1. Dynamics In Supervised Machine Learning Algorithm.

This algorithm's main idea is to map inputs to the desired outputs modeling it through a function. This is regularly used in classification problems, so that the model must learn from a massive data source and, after being trained, outputs a classification for each of the inputs that are given to it.

Supervised algorithms are also defined by their classes, so any data that it collects will be classified into one of them [9]. These algorithms can also be combined with convolutional neural networks, in which there are different layers: an input layer, in which image data can be received, and the convolutional layer, which applies filters to the input and maps the produced outputs. Then there is a pooling layer, that performs a reduction to the spatial dimensions of the convolutional layer, reducing the number of parameters the model works with. Finally, the fully connected layers classify all the data passed to them and produce a final output [10]. A diagram of this process can be seen in Figure 2.

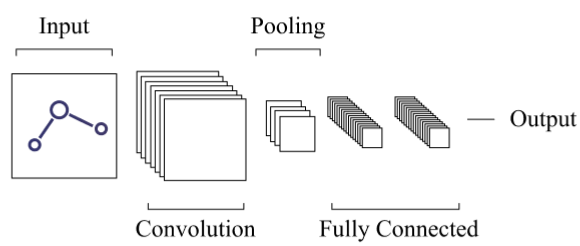


Figure 2. Diagram of a simple five-layered CNN architecture.

Supervised learning, combined with a convolutional neural network was an ideal structure for the machine learning model, considering the team's goal, which was to classify photos of a set of hand-drawn chemical molecules, as there is a standard for drawing them, so the model could use predefined data to predict new image occurrences.

IV. DEVELOPMENT

The development of the application, with its machine learning model, was accomplished using Apple Inc.'s Core ML framework [12]. This framework works together with Create ML, a platform designed for training models that seamlessly integrates into iOS and iPadOS applications. The image classifier within Create ML is designed to match images with predefined labels, efficiently determining the most suitable one for each image, by applying some of the machine learning techniques discussed previously.

The team adopted the Challenge Based Learning (CBL) methodology for this project, fostering collaborative efforts

 ISBN: 978-1-68558-140-4

across its three main phases: Engage, Investigate, and Act. This structured progression provided a systematic approach to the research, enabling effective problem-solving [11].

The first phase (Engage) involved gathering ideas related to the central topic of research, involving the identification of problems that pose challenges. After determining the topic, the group proceeded to choose specific problems and proposed challenges aimed at alleviating or resolving them. The main research question emerged: "How to improve the learning of Chemistry through the use of Machine Learning?".

This central challenge guided the team into the Investigate phase. Now, the focus was on exploring resources to define the problem and propose solutions. Additional guiding questions also emerged, such as: "What are the best methods to optimize student learning?", "Does interactive learning enhance subject comprehension?", and "How can machine learning be effectively used to improve education and be made accessible to students globally?" The concept of an app that uses a camera to identify chemistry molecules stemmed from this phase, with each aspect of the app's development rooted in these formulated questions and emerging answers.

The final phase of the CBL, Act, involved the start of the training of a test model using Create ML. This phase aimed to test the feasibility of the gathered ideas for a solution. It was decided that all photos used to train the model would be taken from regular smartphones and later shared between the team. They consisted of hand-drawn molecules on plain white paper, whiteboards, and ruled paper.

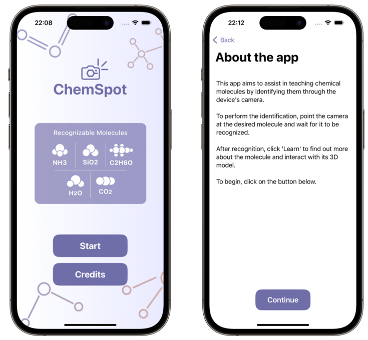
A Core ML model with images of water and silicon dioxide was generated, allowing the team to import it into any project supporting the Core ML framework. This initial model had the intent of evaluating how well the app would distinguish molecules that had similar structures. Our pilot tests exhibited reasonably high accuracy (82%), considering it was trained with just a small set of images. As the dataset expanded, the model's accuracy improved significantly, prompting further training to identify additional molecules.

After acquiring over two thousand images and importing them into Create ML, the final Core ML model was achieved. This model was integrated into an iOS app: ChemSpot, designed and developed to provide comprehensive information about the trained molecules. This app empowers students to identify drawn molecules and explore their properties, formulas, facts, and real-world occurrences.

V. Results

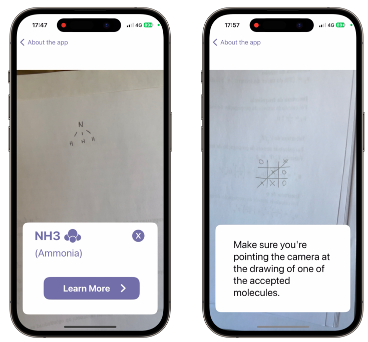
The entire process of training the model, designing, and developing the application was done in a single weekend, with further minor improvements. The CBL methodology was closely followed by the team and the app is currently still in development.

The molecules that the model was trained to recognize are shown right when the application is opened. They are H_2O (water), CO_2 (carbon dioxide), NH_3 (ammonia), SiO_2 (silicon dioxide), and C_2H_6O (ethanol). Those were the chosen molecules for the development of the app given their wide use and occurrence during the teaching of chemistry. Besides those molecules, two additional labels were implemented, so that the model could recognize drawings that were not one of the trained molecules so that the app would not display results.



(a) (b)

Figure 3. (a) Onboarding Screen with List of Recognizable Molecules and (b) Instructions Screen.



(a) (b)

Figure 4. (a) Process of scanning a hand-drawn molecule and (b) state when no molecules are recognized.

(a) (b)

Figure 5. (a) and (b) Learn More page presents the user with relevant information about the scanned molecule.

After clicking "Start" (*Figure 3a*), there is an introduction about the app and its main purpose: recognize hand-drawn molecules and provide useful information about them. This next screen (*Figure 3b*) also offers tips on how to get a better chance of the app recognizing a molecule. The user is then directed to a scanner, using the phone's camera, to identify any of the available molecules and is prompted to learn more about them (Figure 4).

 ISBN: 978-1-68558-140-4

When the user scans a drawing that appears not to be one of the molecules that are recognizable by the app, they are alerted of that, suggesting that they may be pointing to something that the app cannot recognize. When a molecule is recognized and the user clicks on “Learn More”, they are presented with detailed information about the molecule, accompanied by an interactive 3D model of it (Figure 5).

This demo app was created for iOS devices, for testing purposes and so that the app could be run on a phone (Figures 3 to 5), for practicality, or a tablet (Figure 6), where information can be better distributed, due to the larger display.

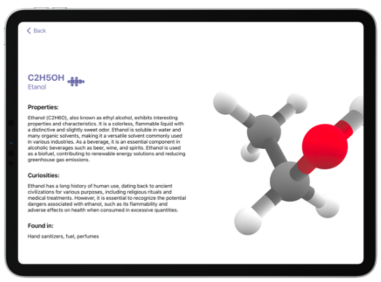


Figure 6. iPad app version of the Learn More page, showcasing a larger interactive 3D model.

Allowing users, especially students, to easily identify hand-drawn molecules and access relevant information about them can be a catalyst for making the study of Chemistry much more interactive and engaging for students. Apps can follow this approach to offer even more resources for students worldwide and create new experiences that aim to improve education.

Although the paper outlines a novel approach and the application seems promising in pilot tests, the full impact and efficacy of the tool will be more accurately gauged with comprehensive user engagement data and a broader dataset for model training.

VI. Conclusion

Technology was a great way to enhance the teaching and learning processes, by providing new ways for students and mentors to search for and interact with new topics and subjects. The pedagogical possibilities it offered were endless, and one of them was bringing interactivity into studying. In the Brazilian scenario, where laboratories were scarce in public schools and the average score for Chemistry in ENEM was below 50%, there was plenty of room for new ideas that aimed to improve the educational system.

This paper showed how machine learning, combined with mobile applications, had the potential to create powerful tools to assist the process of learning, through a demo app that identified hand-drawn molecules and provided information about them, including their properties, curiosities, where they could be found and an interactive 3D model, offering visual aids for the user, which had shown to improve the understanding of abstract subjects greatly.

The app’s model was trained to identify five molecules, but by training it further, more molecules could be implemented. Improvements to the information presented after scanning a molecule, like referencing other learning resources were also a future goal.

Lessons learned from initial failures, such as the challenges of accurately classifying molecules from hand-drawn images and the type of background they were drawn to, reinforced the need for a robust training dataset, which can be seen as future work. The project also highlighted the importance of user interface design that accommodates a wide range of users, from beginner to advanced students.

The use of machine learning for interactive educational tools in chemistry represented a new frontier in pedagogy, especially as it pertained to visual and interactive learning in the Brazilian context.

References

[1] C. F. Alvarenga et al., "Challenges of higher education for public school students: a study at UFLA," Contemporary Thought in Administration Review, vol. 6, no. 1, pp. 55-71, 2012.

[2] L. Massi et al., "Scientific articles as a didactic resource in higher education Chemistry teaching," New Chemistry, vol. 32, pp. 503-510, 2009.

[3] J. A. Pereira, J. F. da Silva Junior, and E. V. da Silva, "Instagram as a collaborative learning tool applied to Chemistry teaching," Debates in Chemistry Teaching Review, vol. 5, no. 1, pp. 119-131, 2019..

[4] I. El Naqa and M. J. Murphy, "What is machine learning?," Springer International Publishing, 2015..

[5] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in 2018 Fourth international conference on computing communication control and automation (ICCUBEA), IEEE, 2018, pp. 1-6.

[6] A. Barros, "Internet reaches 88.1% of students, but 4.1 million from public schools did not have access in 2019," IBGE’s News Agency, 28 April 2021, [retrieved: December, 2023].

[7] BRAZIL. National Institute for Educational Studies and Research Anísio Teixeira (Inep), "Basic Education Census 2019: statistical notes," Brasília, 2020.

[8] K. Demir and E. Akpinar, "The effect of mobile learning application on students’ academic achievement and attitude toward mobile learning," Malaysian Online Journal of Educational Technology, 2018, vol. 6, issue 2.

[9] V. Nasteski, "An overview of the supervised machine learning methods," Horizons. b, vol. 4, pp. 51-62, 2017.

[10] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.

[11] M. Nichols, K. Cator, and M. Torres, "Challenge Based Learner User Guide," Redwood City, CA: Digital Promise, 2016.

[12] Apple Inc., "Core ML | Apple Developer Documentation," [retrieved: December, 2023].

[13] S. J. Obringer and K. Coffey, "Cell phones in American high schools: A national survey," Journal of Technology Studies, vol. 33, no. 1, pp. 41-47, 2007.

[14] M. K. Sarabi and K. A. Gafoor, "Student Perception on Nature of Subjects: Impact on Difficulties in Learning High School Physics, Chemistry and Biology," Online Submission, vol. 8, no. 1, pp. 42-55, 2018.

[15] V. Kukreja and S. Ahuja, "Recognition and classification of mathematical expressions using machine learning and deep learning methods," in 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), IEEE, 2021, pp. 1-5.

Autonomic Computing in Total Achievement of Quality

Joel Bennett, Roy Sterritt
School of Computing
Ulster University
Belfast, Northern Ireland
email: bennett-j8@ulster.ac.uk |
r.sterritt@ulster.ac.uk

***Abstract*—This paper presents a Systemization of Knowledge (SoK) on Autonomic Computing (AC) for Total Quality Management (TQM), i.e. a review of the domain of Quality GxP manufacturing environments considered through the paradigm of Autonomic Computing. The development of autonomic computing concepts and how they are applied currently are discussed. The paper then examines quality systems for each of; manufacturing and automation; product testing validation; data integrity; and supporting IT infrastructure; as pertaining to GxP manufacturing environments, being subject to high levels of regulatory compliance, before concluding with considerations about the need for this self-managing computing paradigm for quality manufacturing, and some avenues of progress identified in the current and future state.**

Keywords-Autonomic Computing; quality system; total quality; achievement; self-x; TQM; MES.

I. Introduction

The objective of this review will be considering the area of Quality manufacturing environments, particularly how those environments are supported by computing systems and how they benefit, or could benefit from the Autonomic Computing (AC) paradigm. Quality, in the context of manufacturing, encompasses a wide range of frameworks, standards and procedures, which include implementation of Good Manufacturing Practice (GMP) [1], Continuous improvement (CI) [2] and for computerised systems supporting the processes, adherence to the Good Automated Manufacturing Practice (GAMP) framework "which aims to safeguard patient safety, product quality and data integrity" [3]. These requirements are commonly found in any setting which requires a high level of regulatory compliance and accountability, such as that found in food, pharmaceutical, or health care.

The goal of these kinds of standards is not always easy to define, but generally is captured under the term Total Quality Achievement (TQA). Standards are updated and there is always an expectation that cases for CI will exist in any organisation, which are found through a combination of internal review and external audits. This is important in ensuring that customer safety standards are met, maintained and kept front and centre.

Underneath the quality activities, there are a range of computer systems and software, from machine automation to product testing, to scheduling and batch release activities, with an overarching Quality Management System (QMS). These environments generate a lot of records and documents, data, data sets and require well defined data retention policies and most often this requires a high level of human effort.

The autonomic question is about how much humans needs to be involved in directly managing systems and how they can be designed beyond this. We can consider first a brief overview of what AC is and then some distinct areas within Quality manufacturing , so as to make some application of it.

This paper presents a Systemization of Knowledge (SoK) on Autonomic Computing for TQM, as such, the first section summarizes AC, then examines quality systems for each of; manufacturing and automation; product testing; validation; data integrity; and supporting IT infrastructure; before concluding.

II. AC & TQM: Systemization of Knowledge (SoK)

This section presents a Systemization of Knowledge (SoK) on Autonomic Computing for TQM.

A. *Autonomic Computing Paradigm*

Having a brief introduction to AC, what it is and sets out to achieve will be useful to understanding where it might fit into the area of Quality systems. The term autonomic is borrowed from the bodies nervous system which governs unconscious functions such as regulating heart rate and temperature, without burdening the conscious area of the brain [4]. As computer and computer supported systems with their software, have become larger, more sophisticated and more interconnected, with growing intranetworks and internetworks, the complexity eventually reaches a level, where the best of systems experts cannot account for all configurations, points of failure and providing timely response to errors in the whole system. The initial recognition of this has its beginnings in IBM, whose Paul Horn, introduced the idea to the National Academy of Engineers at Harvard University in a March 2001 keynote address [5]. IBM envisioned computer systems of systems, with their smallest edge endpoints, up to the largest datacentres, with all of those interconnections as somewhat analogous to the human bodies smallest molecular machines and the bodies signalling equipment, being zoomed out to view entire societies with all their interactions. AC as a paradigm has drawn inspiration from these initial ideas and as IBM and others had predicted, it has become recognised in the computing industry as a necessity to start trying to achieve this goal of systems that are self-managing guided by an autonomic principle.

Following the talk by Paul Horn, IBM released a printed work in October 2001 titled *"Autonomic Computing: IBM's Perspective on the State of Information Technology"* [5]. In this, IBM outlined the problems that AC was seeking to address, the necessity for it and how it might be achieved. The authors present a case, that human progress has always been rooted in the support provided by technology and automation, which frees up human work effort, in order to enable achieving bigger things. However, while computers and the IT industry have supported business and innovation to a certain point, the rising complexity of these systems eventually presents a risk of even reversing these benefits [5, p. 4]. The human effort required to support these very same IT systems

 ISBN: 978-1-68558-140-4

as they expand, rises exponentially. The case is compelling and presents proposals for the capabilities an autonomic system should have, which is provided in 8 main points which are briefly, that an autonomic system should:

1) "Know itself" which is perhaps best summarised by this statement *"a system can't monitor what it doesn't know exists"*. [5, p. 21] This self term has become definitional to AC component descriptions throughout the field.

2) Be able to configure and re-configure itself in response to changing and unpredicted conditions.

3) Not be settled on the current state and always seek to optimise.

4) Be able to perform functions analogous to healing.

5) Be security aware and self-protecting

6) Know its supporting environment and activities, so that it can respond appropriately to them

7) and therefore must not be sealed off and isolated, but must be able to function interdependently facilitated by open standards

8) Constantly anticipate what resources are needed. *[5, pp 21-31]*

This is expanded upon in a further article published by IBM Systems Journal 2003 "The dawning of the autonomic computing era", in which the need for AC is re-stated and then how the industry can begin to adapt by shifting its design objectives from price/performance to instead prioritising "robustness and manageability" and the cost of ownership. [6 p. 7] Here autonomic self-x properties are also elucidated, namely the fundamental self-Configuring, self-Healing, self-Optimising, self-Protecting, which is otherwise captured by the term self-CHOP [6, pp. 8-9].

1) Self-Configuration – is awareness of components and the environment, so as to be able to dynamically re-configure, to automatically integrate new components and adapt. A simple example might be the Plug and Play or hot-swap features of modern hardware and operating systems, that are completely unsupervised after the component has been introduced. This is a distinct development upon the original features of automatic detection, followed by configuration wizards, manual bus assignments, or disk drive rebuilds.

2) Self-Healing – detect and diagnose errors in components, isolate and repair them, or disconnect as necessary. Prevent failures from occurring if possible through component management with a view to maintaining constant availability.

3) Self-Optimising – continual automatic tuning across the systems available resources. This may include some element of built-up knowledge, in order to predict and schedule, as well as the ability to dynamically allocate resources in response to demands.

4) Self-Protecting – Securing system resources via users identification, intrusion detection, secure backup and restore services.

A framework for achievement of this goal, was described in a so-called intelligent control loop as outlined in IBM's work on an autonomic blueprint [7] known as MAPE-K (in Figure 1). Envisioned as an abstraction from the underlying managed resources, an Autonomic Manager (AM), of which there may be several for different specialties, is based upon the MAPE-K loop, though not all of those capabilities need necessarily be used in every situation. AM's could be in a dedicated role, e.g., Self-Configuring, Self-Protecting, while other AM's can occupy a higher level with overall system supervision, described as orchestrating AM's.

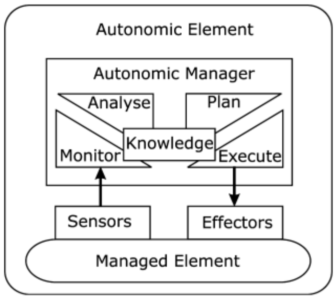


Figure 1. MAPE-K intelligent control loop [8, p. 6]

The make-up of the AM, includes Sensor and Effector interfaces which make one AM available to other AM's and the system components. From sensors, the Monitoring collects and filters data from a resource, which is then Analysed, providing correlations, modelling and predictions so the AM can learn from its environment. In Planning, the AM formulates strategy utilising its formed policies and finally these plans are Executed, whilst remaining open to being updated by new information. The results of these cycles update Knowledge for the purpose of improving outcomes.

With AC implementation as a background, we will now consider a few areas of Quality systems, whether they benefit from autonomic computing currently and where applicable, if there is any future development we might expect.

B. Quality Systems - Manufacturing and Automation

Manufacturing relies upon scheduling and execution controls, usually built upon a Manufacturing Execution System (MES). Typically, an MES system will provide some kind of information about a manufacturing floor and a level of control, for example if certain limits are exceeded, they will be reported to the appropriate receiver [9 p.3]. Computerised automation has greatly enhanced the ability of manufacturers to scale production and improve product quality, but has also increased system complexity. Tasks such as transfer of raw materials whether obtained, retained or disposed, must go through and from approved suppliers. The customers, manufacturer and suppliers often need to audit one another. Every record pertaining to the manufacturing process, including, but not limited to documentation of batch records and product release must be retained and retrievable. Such data integrity quality requirements will be considered later. Product manufacture may, for the most part, be described as automated, but is still heavily supported at almost every level by human activity and decision making. It seems in MES and automation, we can see some parallel, with the problems IBM drew attention to in its early autonomic works on computing.

Autonomic Smart Manufacturing [9] has been proposed for improving upon MES, modelling itself upon the MAPE-K framework, as in Figure 2, using a monitoring phase to collect metrics relevant to the manufacturing process, which are analysed to infer unknown relationships within the environment using machine learning (ML) to make what-if predictions, thus anticipating situations and performing the necessary calibrations. A plan phase would further reflect upon the findings, with a holistic treatment of individual behaviours allowing the system to propose and implement optimisations. In this model, there is still human supervision of the returns by an engineer, which is important to a quality process in terms of accountability, but an autonomic controller maintains optimal parameters as conditions change using a toolkit of built-up models in a knowledgebase to minimise human resource use.

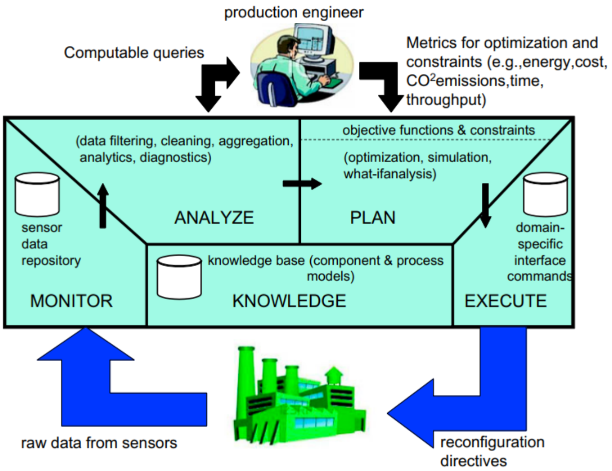


Figure 2. MAPE-K applied to smart manufacturing [9, p. 5]

A further work on "Generative simulation modeling of an Autonomic Manufacturing Execution System (@MES)" [10] noted that traditional MES systems rely on rigid schedules which are shown to be inefficient and ineffective. It proposes a multiagent simulation model where an Order Acceptance autonomic manager has end-to-end knowledge of the shop floor, customer orders and can delegate tasks, while a number of Resource Agents have supervision of specific areas of the MES. The proposal is to implement a loop which incorporates the scheduling and control functions for specific resources, allowing the system to respond dynamically to shop floor requirements (environment) and reconfigure accordingly, whilst also informing the other agents, which will likewise recalibrate. The agents can simulate scenarios (i.e., is the next order feasible?) in conjunction with one another and are able to autonomically optimise routes, responding with suitable planning and execution. A later 2012 paper proposing a selfish multi-agent MES system atop @MES, states one issue remains in that enterprise networking of MES systems remains an open problem [11].

In terms of Quality systems, the concerns around such proposals, may be the validation of the autonomic agents controlling manufacturing processes at a high level and how to ensure accountability and oversight. However, if the autonomic agent has a validated means of justifying its mitigations, then perhaps this concern could be overcome. Validation will also be considered later.

C. *Quality Systems - Product Testing*

Product testing is necessary in any manufacturing process, but the burden of regulatory concerns in quality manufacturing is in many ways higher and must meet the requirements of an internationally recognised standard QMS such as ISO9001. Product testing supports the development of product, as well as forming part of the batch release and CI processes. Chiefly we will consider the typical laboratory setting, where product samples are taken from a batch, whether for production, or from a development cycle. The laboratory comprises a number of systems and instruments forming a testing suite and will of course vary, depending upon the kinds of products being analysed. The instruments utilised in testing must comply with national standards for regular calibration. Failed samples are reported against a batch, so that a determination can be made by quality assurance, investigating whether the sample test failure affects the whole batch, if further testing is required and what is the root cause of the product failures. The results of testing must be retained along with the records of any investigations.

The types of computing systems supporting these operations are typically the Laboratory Information Management Systems (LIMS) supporting clerical activity around recording test results [12], instrumentation control software on an integrated, or external computer and statistical analysis tools. Usually, each test or the days testing is preceded by a sample run to ensure the instrument is operating within defined parameters. Some tests are quite short in duration and others can run for many days. The instrument machinery is largely automated once configured and validated, but experiments are selected manually, including the passing of results and aforementioned calibration.

Laboratory performance is measured by the number of tests performed against erroneous tests performed. LIMS were created to improve the automation of product test data flows and ensure integrity, thereby reducing error and this is the main motivation for adoption of LIMS systems by laboratory management [12 p.2]. While LIMS systems exist that are automatically collecting and approving results, there are also many that involve manual entry of results. Many do incorporate, as with the MES and AM, a part of their control loop which automatically notifies and involves the relevant party when this is necessary. However, very often this is, per machine automation control, based purely on preset tolerance value thresholds.

There does appear to be little work done in the area of autonomic laboratory systems, but we can perhaps identify some autonomic-in-principal elements. There are many safeguards included with laboratory instrumentation, both with a view to safe-guarding results and the instrument itself. Many instruments have awareness of and do not allow operation outside the pre-defined calibration windows. This avoids producing costly invalidated results. Instruments also have internal sensors and diagnostics which prevent operation if a fault is encountered, or if a part has passed its expiration date. Very often, an instrument may have modules for different kinds of experiments and if not needed, the module may be bypassed.

Certainly, an implementation such as the proposed @MES will also need RA's associated with testing in order to simulate requirements. Furthering of autonomic principles in this area, may lead to systems utilising acquired knowledge during analysis to self-adapt. For example, if the current state involves manual investigation and root cause analysis due to test results, autonomic self and environment awareness, perhaps facilitated by other autonomic agents may allow for automatic determination of root cause. An external factor, such as the temperature of the laboratory may be a simple example. Even more desirable, might be autonomic agency preventing invalid tests from being conducted in the first instance and informing the other instruments of current issues. Also, in an autonomic laboratory system of systems, the whole test suite could likewise message other agents that a particular test has successfully completed, allowing subsequent tests on hold for another instrument to proceed, thus closely relating the data generated by those tests.

 ISBN: 978-1-68558-140-4

D. Quality Systems - Validation

The purpose of validation is per GAMP to provide documentary evidence to support a high level of confidence that all parts of a system will work correctly when used [13 p.1]. It comprises various levels of qualification, which describe and contain activities that test the functions a system is supposed to be able to perform. The qualifications include a Design Qualification (DQ) – documenting that all quality aspects of the system have been considered during the design of system, Install Qualification (IQ) – ensures that a system has been installed per its specifications, Operational Qualification (OQ) – as implied, that the listed system functions operates as expected according to the tasks which have been identified, Performance Qualification (PQ) – is evidence that the system works on an ongoing basis in its final setting. Validation applies in a quality setting whether a hardware, or software implementation. In some cases, it may include auditing the vendor of the product to be validated and one reason for this, is that they are supplying some of the documentation that supports the validation – for example, the IQ and OQ.

Security is also an important area of system validation, including that logins work correctly [13 p.3], since it has a direct bearing on the accuracy of records and traceability as per FDA Title 21 CFR Part 11. Data entry validation is also an important aspect, ensuring that data is formatted and saved appropriately, as well as being retrievable [13 p.4]. Evidence for the tests having been carried out may also be required where this is part of the specification.

The creation of design and test validation documents, collection of evidence and need for reviewers and approvers can be a lengthy and time-consuming process. Automation of systems has increased rather than decreased the level of effort required during validation, due to remaining distrust of automated systems, even when automated software testing is considered. Concerns about transparency arise, particularly when increasing automation may come to rely upon black box solutions where the underlying reasoning behind an activity cannot be directly observed. Automated testing and modelling can assist with allaying fears, such as injecting deliberate faults to see how the system handles different scenarios [14] and automated software testing is sometimes employed in validation, but automation can surely only get us so far.

If instead these systems were built autonomically from the ground up, with the autonomic agents forming almost a digital twin of business level hierarchy, which is trusted to provide the same level of assurance for each respective area, as human information gathering and review authority, then this would have the effect of increasing confidence in all of the systems which implement this by virtue of being recognised as self-correcting, self-securing, self-optimising, self-healing. We propose, that it is conceivable, that systems could become self-validating.

E. Quality Systems - Data Integrity

Integral to modern quality systems is Data Integrity (DI), which is data that meets standards of completeness, accuracy and consistency, i.e., the data must be ALCOA, i.e. attributable to a person or persons, legible, contemporaneous to what is being recorded, the original record and accurately recorded [15]. DI also requires consideration for how long records should be retained. On a computer, data can be in raw form, or processed form, but even in the current state of a heavily computerised environment with relatively high levels of automation, the original primary data records are often still paper based. However, as automation moves towards industry 4.0 smart manufacturing, much of the discussion inevitably turns to digitised data and data security.

A large concern in DI, is not just the maintenance of the original record [23 p.3], but the assurance of a validated backup and restore functionality, along with the data retention [23 p.49-50]. Legacy backup solutions were often manual, or even if automatic/scripted are triggered by simple rules within a time window. If a backup is missed, it likely does not run at all, although it may be followed by a notification. Many modern backup solutions do incorporate some autonomic elements, such as those described in "Lifeboat" for IBM, as far back as 2004 [16]. The solution proposes a decentralised peer-to-peer network backup model, with a distributed file system and awareness of disk quotas on each participating system. This eliminates single points of failure and relies on autonomic agency to determine the most appropriate use of available resources. It also discusses the scenarios of a server addition to the system, since availability of clients cannot always be guaranteed, as well as local backups. Many of the concerns with peer availability, have been solved in subsequent technology solutions which implement a Grid computing approach. Sharding algorithms can distribute redundant copies of partial data efficiently across as many peers as available and reconstruct it from x number of peers, way storage systems can reconstruct from parity data across x numbers of disks. A prime example of this in practice with regards to storage are Microsoft's own DFS and BranchCache technologies. The other main autonomic element of the IBM Lifeboat system seemed to be self-configuration in having awareness of new clients by referring to an asset collection database and adding detected clients to the backup automatically. Current backup solutions are relatively self-aware of available bandwidth and the size of data on individual clients, so as to be able to automatically allocate the order and groups of clients to backup queues ensuring that this happens within the provided timeframes. They can also perform self-signing or verification of the backup's integrity using hash checks. Very often the backups are self-optimising and regularly consolidate sets to remove duplication and clean-up to free storage. It seems reasonable to call these elements autonomic developments. However, most backup solutions to the present are still centralised, even if with redundancy. The incremental improvements are welcome, but achieving truly autonomic backup and restore would seem to require something further. Almost all backups still rely on schedules, but in a Quality environment any loss of data, even if rare is not acceptable. Even if the time gap is only a few hours between backups, data lost due to disk failure can affect product release, or manufacturing traceability and will be questioned by an auditor. Simply increasing the schedule frequency may be effective in reducing risk and simple, but is not an autonomic approach and increases system utilisation. The backup solution could instead have an AM, which senses relevant disk transactions and efficiently commits those to the backup storage incrementally. Similarly, being able to restore a failed disk from a backup is not an autonomic solution. An autonomic backup should try to anticipate the fault and ensure the system data is safeguarded and then offline the faulty component. When the fault is addressed, the AM should be able to recover from the fault without user intervention. Perhaps it could also be aware of similar/like systems and offer another suitable systems resource to carry on performing the same function, by making the faulty systems data and functionality available non-destructively. A similar solution was proposed in 2000, albeit using limited computing nodes, that was able to autonomically transfer running applications complete with memory and CPU register content, from faulty nodes to any available working node, which ensured continuity [17]. There has been much progress in the area of thin applications, which can make this level of availability more common. The main issue for a quality environment, is to establish trust of the underlying AM decisions and how to validate data integrity.

Most quality systems rely on a database to store, maintain and retrieve their generated data. Traditional relational databases may have some degree of autonomic design. Many include a self-repair and compact functionality which also serves as a self-optimiser and this is also recognised in an analysis published by IEEE in 2003 [18]. However, the limitations of several popular contemporary DBMS products were also discussed in terms of how they failed to be autonomic and how these products may get there. Under the heading "what is missing?", the article describes the high level of human input needed, the inclusion of data advisers and wizards, rather than self-configuration and the lack of self-optimisation in the form of ensuring the most efficient memory usage with optimal indexing. Likewise, databases tend to include recovery tools and not necessarily the autonomic self-healing property.

Another key component of quality systems data integrity since the late 90's is the Audit Trail. An audit trail provides evidence of actions performed by the system, or in the system. A weakness of the current audit trail implementations, is that, it often doesn't actually influence outcomes, but merely records activities. A security audit trail records that a user/system login took place, or that a particular record was saved by a given logged in user at a given time, but it is often a flat text log file, which doesn't have any actual connection to the potentially affected data records.

A type of database paradigm which seems to address the limitations of traditional database and logging systems in this respect is blockchain technology. The underlying principles of a blockchain are essentially autonomic. A blockchain database consists of a series of blocks, to which any kind of data can be written, and that data is then encrypted by a hash. The entries are both time-stamped and signed by a unique identifier which indicates ownership of the block, even if that identification is anonymous, the transactions are completely transparent on a ledger. Every subsequent entry relies on the hash of the previous block to decrypt itself, forming a chain. It is impossible to update the historical blocks without possessing every single cryptographic hash key, so it is self-protecting. The earliest blockchains were designed around the concept of functioning as currency, with unique identifiers being wallets, but more sophisticated blockchains can integrate so-called smart contracts which are automatically executed on the chain once agreed between parties, as well as hosting digital applications. Underneath the data layer of a blockchain is the concept of the nodes whether full, or partial, which host identical copies of the blockchains data and increase the reliability of the network and its bandwidth, whilst also providing a framework for consensus that block transaction being committed to the database are valid. Blockchains are largely self-configuring, in that they require no user intervention as to their data structure once started and automatically eliminate invalid blocks. They are self-healing because any potential corruption is automatically eliminated by making comparison with other nodes. They can also include apoptotic self-destruction of transactions once these have expired, which automatically releases the storage utilised by the transaction block.

Therefore, given these properties, it isn't surprising that a framework for utilising blockchain in an ISO compliant QMS to achieve Total Quality Management (TQM) has been suggested [19], but it would seem particularly suitable for ensuring contemporaneous association of data records with a digital identification, as per the requirements of DI, without the need for a separate audit trail.

F. Quality Systems - Supporting IT Infrastructure

Finally, although some of the areas above have touched upon the associated IT technologies which support the various quality processes, a brief consideration of how IT infrastructure as a whole supports these systems and how IT has benefited and may yet benefit from autonomicity to support quality seems appropriate. IT Infrastructure refers to all of the components necessary to deliver IT services within an organisation, e.g. equipment, network, software and services, including Internet based services and datacenters [24, p32]. A subset of IT Infrastructure examples are briefly considered below, which have been selected due to suitable existing and potential AC properties.

1) Storage – the dominant centralised server storage paradigm in enterprises of all sizes remains some version of Redundant Array of Independent Disks (RAID), usually level 5 or above. RAID storage systems have been available since at least the early 1990's and above 0 have a self-healing functionality [20, p.8] whether via duplication, or parity, as well as self-configuration in allowing hotswap disk replacement. It seems that due to historicity, these autonomic mechanisms being invisible to the quality process are trusted to perform their data management tasks. This may form a basis for trust in further autonomic developments.

2) Network – most business networks use a tiered star topology utilising centralised switches. Setting aside the increased expenditure of additional cabling and network switches, this is an improvement on the old bus-based networks, since the failure of one cable does not bring down the others segments. However, network switches and ports still often represent single points of failure. There may be a case for Survivable Network Architectures (SNA) as employed in telecommunications [21 p.2] for certain quality operations, particularly automated machine networks in which the network supports communication between machine components and when down causes the whole system to cease operating. Many modern managed switches do offer some autonomic features. Spanning Tree Protocol (STP) prevents physical network loops, by blocking them at Layer-2. Quality of Service (QoS), manages traffic and automatically adjusts network bandwidth allocated to prioritise specific services.

3) Servers – a server historically was usually hardware dedicated to a specific service, or set of services with the needed storage, memory and CPU specification. Typically, in the last 10 years at least, localised servers have become divided into storage arrays, shared via dedicated Storage Area Network (SAN) and separate server hosts, which run the services in virtualisation containers and share their memory and CPU across the respective virtual servers. Virtualisation has incorporated many autonomic elements, including features such as VMWare Fault Tolerance (FT) [22, p.4] which can automatically migrate a running server from one host to another is a fault condition is detected. This kind of self-awareness and corrective action is essentially autonomic. VMWare FT also uses heartbeat monitoring to detect server

 ISBN: 978-1-68558-140-4

crashes, borrowed directly from AC designs. Many physical servers employ heartbeat health monitoring (HBM) to notify IT of any hardware/software issues which are detected. Again, because it is invisible to the end user, these technologies are generally trusted, even if data migration is involved.

4) DataCentres/Cloud – cloud services, hybrid cloud and Software as a Service (SaaS) being hosted in large international datacentres have rapidly gained traction in recent years and many quality supporting applications have begun to transition to, or incorporate cloud elements, as well as an increasing number of standard business applications. Datacentres are essentially by todays standards autonomic powerhouses, with redundancy and the capability to self-configure an entire server loss with zero-touch. However, although telecoms services are generally reliable, the rarity of redundant Internet access required to use these services remains a concern.

III. CONCLUSIONS

In reviewing the state of Quality systems and AC, there seemed to be remarkable parallels between the steps and goals used in TQM and the descriptions of AC processes, such that AC can be envisioned in the various frameworks performing the same functions which are currently requiring a high level of human input. The need for more autonomicity in the various components of the quality system is apparent, but so is the difficulty of overcoming concerns around trust and accountability. There are some ethical concerns around impacts upon job satisfaction, when autonomous systems add autonomic, as it has been noted that the introduction of LIMS "led to an explosion of paperwork" and that automation "usurped" any control/autonomy a worker had and handed it directly to management [12, pp. 1-3].

It was noteworthy, that a lot of literature tends to focus upon the potential for autonomous aspects of systems, perhaps enabled by machine learning and not necessarily the autonomicity of the systems which will allow them to function independently in a trusted way.

Among emerging technologies, Blockchain is an exciting autonomic development which may gain traction in quality environments future state, where ultimate performance is not the highest concern.

Open autonomic standards [6, p.4] will continue to be a way forward and crucial to allowing Quality environments to trust and utilise AC. In the authors' view, the achievement of AC and TQA will mature together.

ACKNOWLEDGEMENTS

This paper was produced as part of COM760 Autonomic Computing & Robotics for Ulster University's MSc. in Artificial Intelligence.

REFERENCES

[1] World Health Organization (WHO), "Good Manufacturing Practices", Available:https://www.who.int/teams/health-product-policy-and-standards/standards-and-specifications/gmp, [Accessed 02.2024].

[2] J. Singh, and H. Singh, "Continuous improvement philosophy – literature review and directions", Benchmarking: An International Journal, Vol. 22 No. 1, pp. 75-119. Available: https://doi.org/10.1108/BIJ-06-2012-0038

[3] ISPE, "GAMP 5 A Risk Based Approach to A Risk-Based Approach to Compliant GxP Compliant GxP Computerized Systems.", 2nd ed., 2022.

[4] J. O. Kephart, and D. M. Chess, "The vision of autonomic computing," in Computer, vol. 36, no. 1, pp. 41-50, Jan. 2003, doi: 10.1109/MC.2003.1160055.

[5] P. Horn, "Autonomic Computing: IBM's Perspective on the State of Information Technology", IBM Presentation, 2001.

[6] A. G. Ganek and T. A. Corbi, "The dawning of the autonomic computing era", in IBM Systems Journal, vol. 42, no. 1, pp. 5-18, 2003, doi: 10.1147/sj.421.0005.

[7] IBM, "An architectural blueprint for autonomic computing." IBM White Paper, 2006.

[8] M. C. Huebscher, and J. McCann, "A survey of autonomic computing—degrees, models, and applications", ACM Comput. Surv., 40, 3, August 2008, DOI:10.1145/1380584.1380585.

[9] D. A. Menascé, M. Krishnamoorthy, and A. Brodsky, "Autonomic smart manufacturing", Journal of Decision Systems, 24:2, 206-224, 2015, DOI:10.1080/12460125.2015.1046714

[10] M. Rolón, and E. Martínez, "Agent-based modeling and simulation of an autonomic manufacturing execution system, Computers in Industry", Volume 63, Issue 1, 2012, pp 53-78, doi: 10.1016/j.compind.2011.10.005.

[11] M. Rolón, and E. Martínez, "Agent learning in autonomic manufacturing execution systems for enterprise networking", Computers & Industrial Engineering, Volume 63, Issue 4, 2012, pp 901-925, DOI:10.1016/j.cie.2012.06.004.

[12] J. E. H. Stafford, "LIMS: an automating or informating technology?", Laboratory Automation & Information Management,Volume 33, Issue 3, 1998, pp 163-168, doi:10.1016/S1381-141X(98)80002-6.

[13] D. Friedli, W. Kappeier, and S. Zimmermann, "Validation of computer systems: Practical testing of a standard LIMS", Pharmaceutica Acta Helvetiae, Volume 72, Issue 6, 1998, pp 343-348, DOI:10.1016/S0031-6865(97)00032-0.

[14] C. Ebert, and M. Weyrich, "Validation of Autonomous Systems", InfoQ, 2021, Available: https://www.infoq.com/articles/validation-autonomous-systems/, [Accessed 02.2024].

[15] H. Alosert, et. al., "Data integrity within the biopharmaceutical sector in the era of Industry 4.0", Biotechnology Journal, 17, 2022, DOI:10.1002/biot.202100609.

[16] T. Bonkenburg, et al. "LifeBoat: An Autonomic Backup and Restore Solution." LISA '04: Proceedings of the 18th USENIX conference on System administration, November 2004, pp 159–170.

[17] R. Mukai, S. Yamada, S. Tanaka, A. Tanaka, M. Kubota, and I. Kogiku, "A networkwide backup system with inter-memory autonomic copy mechanism", Syst. Comp. J., 34: 2003, pp 89-99. DOI:10.1002/scj.1211

[18] S. Elnaffar, W. Powley, D. Benoit, and P. Martin, "Today's DBMSs: how autonomic are they," Proceedings of the 14th International Workshop on Database and Expert Systems Applications,. Prague, Czech Republic, 2003, pp. 651-655, doi: 10.1109/DEXA.2003.1232095.

[19] R. Muruganandham, K. Venkatesh, S. R. Devadasan, and V. Harish "TQM through the integration of blockchain with ISO 9001:2015 standard based quality management system", Total Quality Management & Business Excellence, 34:3-4, pp 291-311, 2023, doi: 10.1080/14783363.2022.2054318

[20] R. J. T. Morris and B. J. Truskowski, "The evolution of storage systems," in IBM Systems Journal, vol. 42, no. 2, pp. 205-217, 2003, doi: 10.1147/sj.422.0205.

[21] R. Sterritt, "Autonomic networks: engineering the self-healing property", Engineering Applications of Artificial Intelligence, Volume 17, Issue 7, 2004, pp 727-739, DOI:10.1016/j.engappai.2004.08.028.

[22] D. J. Scales, M. Nelson, and G. Venkitachalam, "The design of a practical system for fault-tolerant virtual machines", SIGOPS Oper. Syst. Rev. 44, 4, December 2010, pp 30–39. DOI:10.1145/1899928.1899.

[23] Ronolo, S.C., "Assuring Data Integrity towards Regulatory Compliance: A Study on Process Improvement in Data Integrity Compliance of Computerized Systems.", May 2023,

[24] Laan, Sjaak. "IT infrastructure architecture-infrastructure building blocks and concepts second edition". Sjaak Laan, 2012.