UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE TECNOLOGIA E GEOCIÊNCIAS

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

CAIO BEZERRA SOUTO MAIOR

DEVELOPMENT OF MACHINE AND DEEP LEARNING BASED MODELS FOR RISK AND RELIABILITY PROBLEMS

Recife

2020

CAIO BEZERRA SOUTO MAIOR

DEVELOPMENT OF MACHINE AND DEEP LEARNING BASED MODELS FOR RISK AND RELIABILITY PROBLEMS

Doctoral thesis presented to the Programa de Pós-Graduação em Engenharia de Produção to Universidade Federal de Pernambuco for the doctorate degree attainment as part of the requirements of the Engenharia de Produção.

Concentration area: Operations Research.

Advisor: Prof. Dr. Márcio José das Chagas Moura.

Recife

2020

Catalogação na fonte
Bibliotecária Margareth Malta, CRB-4 / 1198

S728d Souto Maior, Caio Bezerra.
Development of machine and deep learning based models for risk and reliability problems / Caio Bezerra Souto Maior. - 2020.
86 folhas, il., gráfs., tabs.

Orientador: Prof. Dr. Márcio José das Chagas Moura.
Tese (Doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2020.
Inclui Referências.
Texto em inglês.

1. Engenharia de Produção. 2. *Machine learning*. 3. *Deep learning*. 4. Tempo de vida útil residual. 5. Detecção de sonolência humana. 6. Monitoramento de equipamento de proteção individual. I. Moura, Márcio José das Chagas (Orientador). II. Título.

UFPE

658.5 CDD (22. ed.) BCTG/2020-152

CAIO BEZERRA SOUTO MAIOR

DEVELOPMENT OF MACHINE AND DEEP LEARNING BASED MODELS FOR RISK AND RELIABILITY PROBLEMS

Doctoral thesis presented to the Programa de Pós-Graduação em Engenharia de Produção to Universidade Federal de Pernambuco for the doctorate degree attainment as part of the requirements of the Engenharia de Produção.

Approved in: 19 / 02 / 2020.

EXAMINATION BOARD

Prof. Dr. Márcio José das Chagas Moura (Advisor)
Universidade Federal de Pernambuco

Prof. Dr. Cristiano Alexandre Virgínio Cavalcante (Internal Examiner)
Universidade Federal de Pernambuco

Profª. Drª. Isis Didier Lins (Internal Examiner)
Universidade Federal de Pernambuco

Prof. Dr. Paulo Fernando Ferreira Frutuoso e Melo (External Examiner)
Universidade Federal do Rio de Janeiro

Prof. Dr. Marcelo Ramos Martins (External Examiner)
Universidade de São Paulo

ACKNOWLEDGEMENTS

Gostaria de demonstrar meus profundos e sinceros agradecimentos a:

- Meus pais, Leonardo e Lúcia, por se esforçarem ao máximo para me garantir uma boa formação e me encorajar, tanto nos estudos como nos demais aspectos. Agradeço a eles o apoio incondicional em toda a minha vida. Minha irmã, Thalita, que sempre esteve ao meu lado e buscou me ajudar da maneira que ela podia;
- Minha vó Maria, por seu amparo em qualquer questão relacionada não só à educação, mas também à vida. Minha vó Rosa, por sua vibração e alegria com as conquistas realizadas nesses anos;
- Mayra, que acompanhou meus últimos passos nessa tese, sendo compreensiva e me apoiando em vários momentos;
- Todos os meus amigos do CEERMA, que sempre forneceram um ambiente agradável e cooperativo para meu desenvolvimento acadêmico e intelectual. Em especial, gostaria de agradecer a João, Lucas e July, que trabalharam junto comigo e me auxiliaram em vários pontos ao longo dessa desta tese;
- Professor Márcio Moura, meu professor orientador, pela ajuda nos vários desafios e incertezas ao longo do doutorado, estimulando-me a dar sempre o melhor de mim e lutar por todas as oportunidades. Agradeço às suas observações que me fizeram crescer não só na academia, mas também pessoalmente;
- Professora Isis Lins, pela sua disponibilidade, suporte e importantes sugestões de melhorias para esta tese, além de ser sempre uma palavra positiva nas dificuldades;
- PPGEP e CAPES (Código de Financiamento 001) por terem disponibilizado o conhecimento e o apoio financeiro durante meu doutorado acadêmico.

"By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it."

Eliezer Yudkowsky

ABSTRACT

Artificial intelligence-based algorithms have evolved dramatically over the last couple of decades. Specifically, Machine Learning (ML) and Deep Learning (DL) models have emerged as solutions for many tasks previously unreachable, bringing innovation to the industry, with autonomous driving cars and smart houses, and revolutionizing the society with applications going from movie recommendation to medical diagnosis. In this context, this thesis proposes and brings discussion to ML and DL methodologies successfully developed for three distinct problems in applications related to risk and reliability engineering. In the first, a drowsiness detection model is developed to avoid accidents caused by inattention in the context of human reliability. The second problem deals with estimations of remaining useful life of bearings in the prognostic and health management context. In the last problem, a system to detect usage of personal protective equipment in the context to support safety monitoring is presented. In ML methodologies, support vector machines are used, while convolutional neural networks are applied to DL models. Considering the availability and accessibility of datasets, the obtained results demonstrate adequation of methodologies as tools to provide valuable information to support decisions.

Keywords: Machine learning. Deep learning. Human drowsiness detection. Remaining useful life. Personal protection equipment monitoring.

RESUMO

Algoritmos baseados em inteligência artificial evoluíram drasticamente ao longo das últimas décadas. Especificamente, modelos de *Machine Learning* (ML) e *Deep Learning* (DL) surgiram como soluções para muitas tarefas anteriormente inacessíveis, trazendo inovações à indústria, com criação de carros autônomos e *smart houses,* e revolucionando a sociedade, com aplicações indo desde recomendações de filmes a diagnósticos médicos. Neste contexto, esta tese desenvolve e discute metodologias de ML e DL empregadas com sucesso em três cenários distintos em aplicações relacionadas à engenharia de confiabilidade e risco. A primeira aplicação visa desenvolver um modelo de detecção de sonolência para evitar acidentes causados por desatenção no contexto da confiabilidade humana. O segundo problema trata das estimativas de vida útil remanescente de rolamentos no contexto de *prognostic and health management*. No último problema, um sistema para detectar o uso de equipamentos de proteção individual é apresentado como suporte no contexto de monitoramento de segurança. Nas metodologias ML, *support vector machines* são usadas, enquanto redes neurais convolucionais são aplicadas em modelos de DL. Considerando a disponibilidade e acessibilidade dos conjuntos de dados, os resultados obtidos demonstram a adequação de tais metodologias como ferramentas para fornecimento de informações valiosas para o suporte às decisões.

Palavras-chave: *Machine learning. Deep learning*. Tempo de vida útil residual. Detecção de sonolência humana. Monitoramento de equipamento de proteção individual.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI	Artificial Intelligence
APE	Absolute Percentage Errors
CEEMD	Complete Ensemble Empirical Mode Decomposition
CNN	Convolutional Neural Networks
CV	Computer Vision
DL	Deep Learning
DROZY	ULg multimodality drowsiness data-base
EAR	Eye Aspect Ratio
EEG	Electroencephalogram
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EOG	Electrooculogram
EPM	Electronic Performance Monitoring
FPS	Frames per Second
GPU	Graphics Processing Unit
HHT	Hilbert Huang Transform
HOG	Histograms of Oriented Gradient
ILO	International Labor Organization
IMF	Intrinsic Mode Functions
KKT	Karush-Kuhn-Tucker
KSS	Karolinska sleepiness scale
LBP	Local Binary Patterns
ML	Machine Learning
MLP	Multilayer perceptron
NN	Neural Networks
PERCLOS	Percentage of Eyelid Closure
PHM	Prognostic and Health Management

PPE	Personal Protective Equipment
PSO	Particle Swarm Optimization
PVT	Psychomotor Vigilance Tests
RT	Reaction Time
RUL	Remaining Useful Life
STFT	Short-time Fourier Transform
SVM	Support Vector Machine
TBI	Traumatic Brain Injury
WT	Wavelet Transform
YOLO	You Only Look Once

CONTENT

1 INTRODUCTION

In today's competitive business environment, industrial processes need to achieve high flexibility and efficiency, which require companies to deal with data issues of rapid decision-making for improved productivity. Industrial automation is computerizing devices to complete manufacturing tasks with a limited human involvement producing goods quickly and more accurately (AL-FUQAHA *et al.*, 2015).

Transformation from current process in which operators control machines, managers design logistic schedules and machines only perform the assigned tasks, into more intelligent structures requires further advancement in the science by tackling several issues (LEE; KAO; YANG, 2014). According to Lasi *et al.* (2014), the combination of future-oriented technologies in the field of smart objects (machines and products) seems to result in a new fundamental paradigm shift in industrial production, which is the basis of an advanced digitalization within industries. The great potential for revolutionizing all aspects in society relies on the growing body of information hidden in the unprecedented volumes of non-traditional data, which requires both the development of advanced technologies and interdisciplinary teams working in close collaboration (CHEN *et al.*, 2014).

In this context, the term 'Industry 4.0' collectively refers to a wide range of technological concepts in which manufactures would: (1) be completely equipped with sensors, actors, and autonomous controlled systems to monitor production and safety; (2) process parameters of components underlying degradation recorded digitally to provide real condition of the system; and (3) design new manufacturing systems to follow and alert human safety and needs instead of the reverse (LASI *et al.*, 2014).

The goals of Industry 4.0 are to achieve a higher level of operational efficiency and productivity, as well as a higher level of automatization (THAMES; SCHAEFER, 2016). To that end, it should combine the smart objects with big data analytics (WANG *et al.*, 2016).

Almost all disciplines and research areas, including engineering, computer science, business, and medicine, are currently deeply involved in this spreading computational culture of big data due to its broad reach of potential influence in multiple disciplines (BOYD; CRAWFORD, 2012). In recent years, sensors have become cheaper and readily available for industries (SAN MARTIN *et al.*, 2019). Specifically, the massive amount of industrial data generated in these manufacturing environment provides new possibilities for further improvement of reliability and efficiency (REN *et al.*, 2017).

1.1 MOTIVATION AND JUSTIFICATION

Generally, reliability and risk analysis involves the analysis of the data to quantify a prediction and/or to make a decision, which can be, for instance, the probability of failure of one system, remaining useful life of a structure, or trading strategies. Therefore, whether the predictions are acceptable is heavily dependent on how much useful information can be extracted from the original data and how to conduct the analysis (JIANG *et al.*, 2017). Thus, the big data wave only matters if it is useful.

In other words, if not accurately and completely understandable, big data could also bring problems in data quality and data usage. Many manufacturing systems are not ready to manage big data due to the lack of analytic tools (LEE; KAO; YANG, 2014). Analysis based on data with errors results in biased or wrong conclusions and it is necessary to have a deeper understanding of the quality issues of big data and its consequent problems (LIU *et al.*, 2016).

Therefore, new types of advanced manufacturing and industrial processes involving machine-to-human collaboration and symbiotic product realization are emerging (THAMES; SCHAEFER, 2016) and the development of algorithms for dealing with data is one of the major challenges in Industry 4.0 (LU, 2017).

Artificial Intelligence (AI) algorithms are transforming large segments of the economy with expectations created about the intrusion of algorithmic machines into aspects of life previously dependent on human judgment. With the increase in complexity and multiplicity of problems in real world, decisions have been taking place in environments in which the consequences of possible actions are not accurately known and decision making may be uncertain, leading to an unclear future state of the system.

Hence, advanced tools to support decision in economic, environmental, safety and technical contexts have emerged and AI predictive algorithms – such as Machine Learning (ML) and Deep Learning (DL) - are the keys to unlocking the data that can precisely inform real-time decisions (CHEN; ASCH, 2017).

The effect of these learning approaches can already be seen in products and services, including medical diagnoses, product marketing, recognition of spoken language and self-driving cars as well as in industry-related problems, such as consumer services, fault diagnosis in complex systems, control of logistics chains and recognition of human patterns (JORDAN; MITCHELL, 2015). It can be far easier to train a system by showing it examples of desired input-output behavior than to program it manually by anticipating the desired response for all possible inputs and their ability to deliver impressive forecasting power and speed are transforming many facets of society (COGLIANESE; LEHR, 2016).

Research on real-time systems to interpret acquired data and utilize them to make critical decisions advances significantly in areas that constitute a large portion of overhead costs in many industries, such as reliability, maintenance and safety management (Tamilselvan and Wang 2013, Teizer 2015). Therefore, this thesis presents development and discussion of modern ML and DL algorithms in the contexts of risk and reliability engineering.

1.2 OBJECTIVES

The objectives can be identified by general objective and specific objectives.

1.2.1 General Objective

The main purpose of this thesis is to develop and provide discussion of new technologies and AI-based methodologies to monitor key parameters in distinct risk and reliability contexts. Different Machine (ML) and Deep Learning (DL) models are created for three contexts: (i) drowsiness detection in human reliability context; (ii) Remaining Useful Life (RUL) estimations in the context of Prognostic and Health Management (PHM); and (iii) personal protection equipment detection in safety monitoring.

1.2.2 Specific Objectives

In order to achieve the general objective, some specific targets are defined:

a) General understanding and brief review of ML and DL algorithms, focused on Support Vector Machines and Convolutional Neural Network, respectively;
b) Development of ML model to provide real time monitoring of alertness. Analysis and validation of the model on a real data base of drowsiness detection;
c) Creation of ML and DL approaches for RUL predictions. Analysis and validation of model on a real data base of degradation bearing;
d) Adaptation of a pre-trained DL object detection model to identify PPE detection in real time. Application of ML model to verify if PPE is correctly worn.

1.3 OUTLINE OF THE THESIS

Besides this introduction chapter, this thesis has five additional chapters briefly described as follows:

a) **Chapter 2**: Presents the concise theoretical background about ML and DL with emphasis on the specific algorithms used in the application problems (i.e. Support Vector Machine (SVM) and Convolutional Neural Networks (CNN));
b) **Chapter 3**: Provides the first application problem, in which ML is used in the context human reliability. Specifically, a real-time drowsiness detection model is created to warn about human attention;
c) **Chapter 4**: Provides the second application problem, in which ML and DL are used in the context of Prognostic and Health Management (PHM). Specifically, predictions of Remaining Useful Life (RUL) are made to support maintenance policies;
d) **Chapter 5**: Provides the third application problem, in which DL is used in the context safety monitoring. Specifically, a real-time Personal Protection Equipment (PPE) algorithm is created to monitor safety in industrial environments;
e) **Chapter 6**: Contains a summary of the main aspects and results presented in this thesis along with propositions for future work.

In the three application (i.e., Chapter 3, Chapter 4 and Chapter 5), several techniques are applied aiming to develop a suitable model for each specific problem. In Table 1, a brief presentation of the core techniques, which are further explained in Chapter 2, and the support techniques, explained in its respective chapter, applied on this thesis is displayed.

Table 1 – Core and support techniques applied in this thesis

Chapter	Title	Core Techniques	Support Techniques
3	Development of real-time system for autonomous drowsiness detection using eye aspect ratio	Support Vector Machine;	Histograms of Oriented Gradient;
4	Methodologies for estimation of remaining useful life from vibration signal	Support Vector Machines; Convolutional Neural Networks;	Empirical Mode Decomposition; Short-time Fourier Transform; Wavelet Transform.
5	Construction of personal protective equipment detector using video images	Convolutional Neural Networks. Support Vector Machines;	-

Source: The Author (2020)

2 THEORETICAL BACKGROUND

‘‘I am convinced that the crux of the problem of learning is recognizing relationships and being able to use them’’ - Christopher Strachey in a letter to Alan Turing, 1954 (NICKEL *et al.*, 2016).

This chapter presents definitions and explanations about the fundamental topic of this work: Machine Learning, Support Vector Machines, Deep Learning and Convolutional Neural Network. The applications provided in sections 3.3, 4.2 and 5.3 use these concepts.

2.1 MACHINE LEARNING

The central goal of Artificial Intelligence (AI) is to provide a set of algorithms and techniques that can automatically and efficiently solve problems that humans perform intuitively. Machine Learning (ML) is a field of computational learning in AI that looks for the automated detection of meaningful patterns in data and to solve problems which are impossible, or impractical, to be represented by explicit algorithms (BEN-DAVID; SHALEV-SHWARTZ, 2014). While AI embodies a large, diverse set of work related to automatic machine reasoning (knowledge, inference, planning), ML subfield tends to be specifically interested in pattern recognition and learning from data.

Basically, the learning problem could be defined according to its output as (i) classification problems, in which the output assumes discrete values that represent categories; and (ii) regression problems, in which the output is real-valued and its relation with the input is given by a function. Essentially, there are three characteristics in which ML has achieved successful results: (1) a pattern exists; (2) it is not possible (or viable) to mathematically represent it; (3) there exists data relating variables.

The performance of ML algorithms depends heavily on the representation of the data they are given (BENGIO; COURVILLE; VINCENT, 2013). Each piece of information included in the representation is known as a feature. Traditionally, a feature aims to extract useful information presented in the data by means of a high-level representation of the raw data while it reduces the dimension of the problem to be solved, achieving better performance in generalization for unknown data. Many AI tasks can be solved by designing the right set of features to extract for that particular problem, then providing these features to a simple ML algorithm. However, it is often difficult to know what features should be extracted. Manually designing features for a complex task requires a great deal of human time and effort, which can consume greater resources of an entire community of researchers for pre-processing (GOODFELLOW; BENGIO; COURVILLE, 2016).

ML methods could be broadly divided in: (i) supervised learning, (ii) unsupervised learning; and (iii) reinforcement learning. In supervised learning, each pattern is a pair, which includes an input object and a desired output value. In contrast, unsupervised learning models do not need an “expert” intervention and the model is able to find hidden structures in its inputs without knowledge of outputs (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006). In reinforcement learning, algorithms try to find the best ways to earn the greatest reward (MNIH *et al.*, 2016).

The most successful application of ML is, definitely, seen in supervised problems. ML models find relations between inputs and outputs, which allows its usage in context such as data mining, pattern recognition and forecasting problems, for instance. These contexts are particularly interesting because one has to work with big datasets and the task of preprocessing, data preparation and/or prediction can be undertaken by ML models (VOYANT *et al.*, 2017).

Considering the class of supervised learning, well-known methods are generalized linear models, Neural Networks (NN), Naive Bayes, Random Forest (RF) and Support Vector Machines (SVM) (VOYANT *et al.*, 2017). Conversely to most of ML methods, such as NN and RF, that may be trapped on local minima, the SVM resolution entails a convex and quadratic optimization problem in which the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient to provide a global optimum (MAIOR; MOURA; LINS, 2019). Therefore, details SVM are described in section 2.2

2.2 SUPPORT VECTOR MACHINE

According to Wang (2005), Support Vector Machine (SVM) is as a supervised learning method which aims at creating a mapping function between an input vector $\boldsymbol{x}$ and an output scalar y based on the training data set $D = \{(\boldsymbol{x_1}, y_1), \ldots (\boldsymbol{x_m}, y_m)\}$. The objective is to find the function $f(\boldsymbol{x})$ with the smallest penalization with respect to the deviation from the real data and, at the same time, as flat as possible.

SVM, firstly proposed by Vladimir Vapnik, is based on the principle of the structural risk minimization, with its concepts built on the Statistical Learning Theory (VAPNIK, 2000). This means to solve a convex and quadratic optimization problem in which the Karush-Kuhn-Tucker (KKT) condition are necessary and sufficient conditions to guarantee a global optimum.

The goal is not to look for the perfect alignment between the function $f(\boldsymbol{x})$ and D, but the best representation for the mapping (i.e. a trade-off between the data fitness and the generalization ability to predict new data), not requiring any assumption or previous information about neither the function behavior nor the relation between input and output. The regression hyperplane equation is represented by:

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b \tag{1}$$

with $\boldsymbol{x}$ expressing the input data, and $\boldsymbol{w}^{\boldsymbol{T}}$ and b the coefficients of the model. To estimate its values, it is necessary to optimize (i.e. minimize) the regularized risk function provided in Equation ((2):

$$\min_{\omega,b} C\frac{1}{m}\sum_{i=1}^{m}\psi_{\varepsilon}(y_i, f_i) + \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} \tag{2}$$

in which

$$\psi_{\varepsilon}(y_i, f_i) = \begin{cases} |y_i - f_i| - \varepsilon & if \quad |y_i - f_i| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where y_i is the i-th real output (i.e. the original data) while f_i is the i-th estimated value. Equation (3) is known as the Vapnik's ε-insensitive loss function, which implies a non-penalization when the points are inside a tube with radius ε. Hence, ε measures the performance in the training process related to the first term of Equation ((2). The second term of the same equation is used as a smoothness function of $f(\boldsymbol{x})$ and is related to the machine's capacity of generalization represented by $\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}$. Yet, C is a trade-off for penalization between the empirical risk and the model smoothness.

In addition, the problem could be formulated using the primal-dual relation, which states that the solution from the dual problem is also solution for the primal one. In practice, the dual problem is the one actually solved and, from the KKT conditions, a global solution is achieved. For more information related with the primal-dual problem, see Wright (2011). Hence, the optimal estimated regression function $f(\boldsymbol{x})$ is obtained in terms of the dual problem from Equation (1) as follows:

$$f(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\,\boldsymbol{x}_i^T\boldsymbol{x} + b \tag{4}$$

where α_i and α_i^* are the dual Lagrange multipliers. To solve the linear regression, it is necessary to calculate the dot products, $\boldsymbol{x}_i^T\boldsymbol{x}$, $i = 1, 2, \dots, m$.

To generalize and deal with non-linearity, mapping functions $\Phi(x)$ are applied and the dot product, $\Phi^T(x)\Phi(x)$, is solved in higher dimension. However, in addition to the selection of a proper mapping function, the explicit calculation of the dot products with mapped input vectors can be computationally expensive. In this way, dot products are replaced by kernel functions $K(\boldsymbol{x}_i, \boldsymbol{x}),, i = 1, 2, \dots, m$, defined in the input space that implicitly map $\boldsymbol{x}$ into a higher dimension. One can see an interesting tutorial about kernel methods in Jäkel, Schölkopf & Wichmann (2007). Therefore Equation (4) becomes Equation (5):

$$f(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{5}$$

A common kernel function is the gaussian Radial Basis Function (RBF), which is expressed by $K(x_i, x_j) = \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right)$, where γ is also a model parameter. One of several advantages of RBF over others kernel functions is to provide great flexibility requiring just one parameter (Lins *et al.* 2013). In this thesis, SVM models are explored in section 4.2.1, as well as in section 3.3. Also, SVM is used for a specific task in section 5.3.2.

2.3 DEEP LEARNING

Conventional ML techniques, such as SVM, have already been successfully applied on a diversity of reliability problems (e.g. Rocco S. and Zio 2007; Moura *et al.* 2011; Souto Maior *et al.* 2016). Even so, Lecun *et al.* (2015) state that these common approaches have limitation on their ability to process natural data in their raw form and alternatives, such as Deep Leaning (DL), should be studied. Recent advances in deep neural networks, in which several layers of nodes are used to build up progressively more abstract representations of the data, have made it possible for artificial neural networks to learn concepts such as object categories directly from raw sensory data (MNIH *et al.*, 2015)

The current success of DL is directly related with the spread of cheap, multi-processor graphics cards or Graphics Processing Units (GPUs). GPUs are widely used for video games, inserted in a huge and competitive market that has driven down hardware prices. GPUs provide fast matrix and vector multiplications required not only for convincing virtual realities but also for training networks, where they can speed up learning by a factor of 50 and more

(SCHMIDHUBER, 2015). Hence, GPU computing allows increase the speed and decrease the training time when creating a DL model (VERSTRAETE *et al.*, 2017).

Inspired by biological observations on human brain mechanisms for processing of natural signals, DL have been on the spotlight of the academic community (CHEN *et al.*, 2014). It is a specific field that attempts to learn high-level abstractions in data by utilizing hierarchical architectures (GUO; CHEN; SHEN, 2016). In contrast to most conventional learning methods, that considered using shallow-structured learning architectures, DL uses supervised and/or unsupervised strategies to automatically learn hierarchical representations in deep architectures (RANZATO; BOUREAU; LECUN, 2008). DL attracted wide-spread attention outperforming alternative ML methods since 2009, winning many official international pattern recognition competitions and achieving the first superhuman visual pattern recognition results in limited domains (CIREŞAN *et al.*, 2012). It has presented better performance in a variety of problems such as image classification, natural sentence classification, and image segmentation than previous methods based on shallow architectures.

There is no consensus on which problem size divides shallow learning (standards ML) and DL and discussions with experts have not yet yielded a conclusive answer. However, a notion is to consider networks with depth > 3 as DL once problems of depth > 10 are considered very DL (SCHMIDHUBER, 2015).

Even though most DL techniques are capable of automatic feature extraction, great care must be taken when choosing which technique to use when dealing with a specific task (COFRE-MARTEL *et al.*, 2019). Within these techniques, Convolutional Neural Networks (CNN) have proven to be better than to deep neural networks at obtaining a representation of the input data involving grid type data such as images or matrices. Details about this method are described in section 2.4.

2.4 CONVOLUTIONAL NEURAL NETWORKS

A standard artificial Neural Network (NN) is a class of ML algorithms inspired by the structure and function of the brain and consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment while other neurons get activated through weighted connections from previously active neurons. Shallow NN-like models with few stages have been around for many decades (JAIN; MAO; MOHIUDDIN, 1996). For further information of NN, see Cheng and Titterington (1994).

Based on useful insights, architecture and results previously achieved by NN, the majority of DL models focus on training networks with many more hidden layers that are capable of hierarchal learning where simple concepts are learned in the lower layers and more abstract patterns in the higher layers of the network.

Particularly, an operator that is especially interesting for high-dimensional data, such as images and time-series data, is convolution. Indeed, CNN, a dominant architecture of DL for image classification, can rival human accuracies in many tasks (VERSTRAETE *et al.*, 2017). CNN uses hierarchical layers of tiled convolutional filters to mimic the effects of human receptive fields - on feedforward processing in early visual cortex - thereby exploiting the local spatial correlations present in images, and building in robustness to natural transformations such as changes of viewpoint or scale (MNIH *et al.*, 2015).

Conversely to fully connected layers, in a convolutional setting, each input neuron is not connected to each output neuron in the next layer, but is divided into locally connected segments instead (LANGKVIST; KARLSSON; LOUTFI, 2014). CNNs encompass special kinds of NN that include the feature extractor within its training process. The fact that CNNs turned the manual feature extraction design into an automated process is its primary feature and advantage. In an elementary view, CNN is a NN that uses a convolution operation in place of general matrix multiplication (KHAN; YAIRI, 2018).

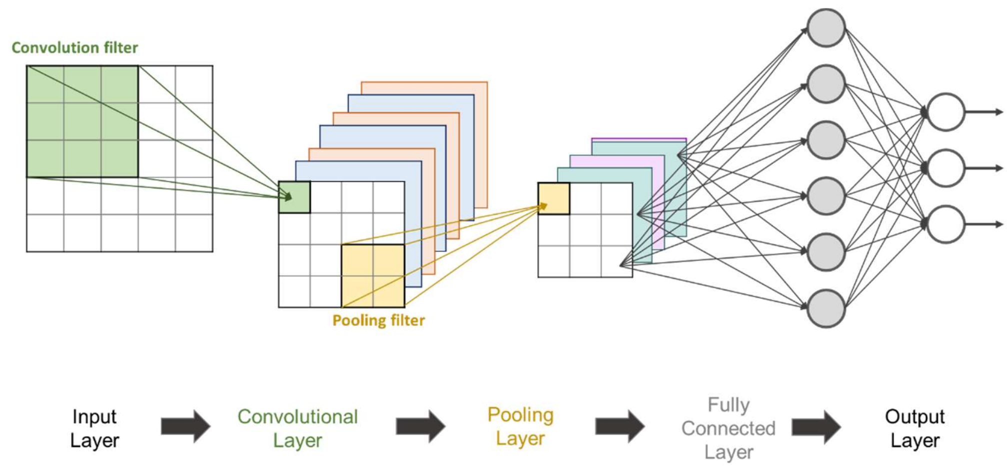
Even though the best achieved results comes for image representation, CNN could handle data that are in the form of different types of multiple arrays: 1D for signals and sequences (e.g. time series); 2D for images (e.g. videos); and 3D (e.g. volumetric images) (LECUN; BENGIO; HINTON, 2015). A typical CNN is composed of many layers of hierarchy with some layers for feature representations (or feature maps) and others as a type of conventional neural networks for classification (CHEN *et al.*, 2014).

The core layers of CNN, convolutional layers, creates a feature maps of the signal analyzed (e.g. an image) connecting with the previous layer with a set of weights called filter bank. Normally, most of architecture starts with convolutional layers to organize the units in feature maps and subsampling (pooling) layers to merge similar features into one feature. Subsampling layers reduce the sizes of proceeding layers by averaging pixels within a small neighbourhood (or by max-pooling) and is normally performed between two convolutional layers. Pooling also reduces the output dimensionality but keeps the most salient information.

After multiple stacks of these layers are completed, the output can be fed into the final stage of the CNN, a fully connected feed forward NN (i.e. multilayer perceptron (MLP)). The outputs of the final pooling layer are used as an input to map labels provided for the data (VERSTRAETE *et al.*, 2017).

The basic structure for CNNs is represented as follows Figure 1: Convolutional – Pooling – Fully Connected Layers. However, deeper architectures including sequences of Convolution-Pooling layers as well as layers for normalization and dropout, are commonly used (LIU *et al.*, 2018).

Figure 1 - Generic two-dimensional CNN



In many applications, CNNs are currently responsible for pushing the state-of-the-art forward in unlikely related fields from speaker recognition (NAGRANI *et al.*, 2020) and person identification (LI; JIANG; HWANG, 2020) to website categorization for e-commerce (BRUNI; BIANCHI, 2020), for example. For an interesting review of the history of NN and DL, see Goodfellow, Bengio, and Courville (2016). In this thesis, CNN models are explored in section 4.2.2 as well as in section 5.3.

3 DEVELOPMENT OF REAL-TIME SYSTEM FOR AUTONOMOUS DROWSINESS DETECTION USING EYE ASPECT RATIO

Some results discussed in this chapter have been published on the Probabilistic Safety Assessment & Management (Los Angeles) in 2018 and are under revision on the journal Expert System with Application in 2019.

3.1 CONTEXT

Nowadays, numerous interventions have been made by governments and industries around the world to improve process safety, with human reliability focusing its attention on risks generated by human errors with high impact of loss involved (BHAVSAR; SRINIVASAN; SRINIVASAN, 2017). Specially, consequences of performance failure by operators in safety-critical task scenarios had increased concerns and drove important research since an inattention or distraction could negatively affect the entire system, including the integrity of the people on it (MAIOR *et al.*, 2018).

Despite the highly reliable equipment, complex safety management regimes and modern automation and control system employed in industries, human operators still have a central role in the execution of complex tasks in which cognitive functions (e.g. working memory, attention, information processing speed) strongly influence their performance, particularly in emergency situations. In addition, high demands in life outside of work – as sleep disturbances and stress – rise as a challenge that may not allow humans for responding effectively to their tasks. Indeed, even temporary failure of cognitive and mental capacities can lead to serious consequences, especially when accurate and immediate response is required (ANSIAU *et al.*, 2008). Hence, understanding the role human errors play in previous accidents is a key information to avoid happening again in the future (RAMOS *et al.*, 2017).

Several studies on alertness have already confirmed that, despite genuine intentions, few watch-keepers remain unwaveringly vigilant while engaged in monotonous monitoring tasks (MAKEIG; INLOW, 1993). Often, the cause of distraction could be related to some kind of tiredness such as mental fatigue and/or drowsiness. The distinction between both is that the former does not rapidly fluctuate over periods of a few seconds, while the latter does. In this way, the challenge remains to detect signals of drowsiness as promptly as the operator (user) starts to present its evidence.

Recognizing human activities from video is a remarkable applications of Computer Vision (CV) (TURAGA *et al.*, 2008). Thus, it is possible to monitor performance of operator to ensure the correct execution of tasks, extracting fatigue related characteristics in real-time by using image/video processing (Vural and Akgul 2009, Rani, Subhashree, and Devi 2016). An important research domain in CV is object detection (BABENKO; MING-HSUAN YANG; BELONGIE, 2011), which has already been applied to different contexts such as surveillance and security (GONZÁLEZ GARCÍA *et al.*, 2017), industrial control (MARTYNENKO, 2017), robotics (HADDADIN, 2014), among others. Indeed, the same concept is also used to track the human body (e.g. eye, mouth, arms) (LI *et al.*, 2013b), recognize its shape and movement (MOESLUND; HILTON; KRÜGER, 2006), create diagnosis (WANG *et al.*, 2010), and infer future behavior (ZELINSKY; PENG; SAMARAS, 2013).

In this chapter, CV tools are used to extract and identify signals of abnormal states (i.e. drowsiness) observed in operator images/videos sensors. The goal is not to detect aspects and status after the failure happens, but proactively identify behaviors performed by the operator before it actually takes place, avoiding all disturbance caused by the error.

Specifically, eye blinks detection plays an important role in systems that monitor human operator's vigilance (Fernandez *et al.* 2016, Salehian and Far 2015). In this context, blink analysis has been object of study applied to memory performance (ZHANG *et al.*, 2015), reflex reaction (ANOKHIN *et al.*, 2003) and drowsiness detection (Schleicher *et al.* 2008, Verwey and Zaidel 2000). However, most works required specialized and highly intrusive machinery to collect data (e.g. biological signals such as electroencephalogram (EEG) and/or electrooculogram (EOG) (ROY *et al.*, 2016). Others come with the CV approach, yet commonly still demands high computational costs/heavy hardware (KIM *et al.*, 2017), having limitations related with user's mobility and camera position (CAVUSCULU; YETIK; YEGINER, 2017), and/or requiring special devices to process (DEMENTYEV; HOLZ, 2017).

Therefore, an adaptable method is created to detect drowsiness directly from an ordinary web cam is valuable. In order to evaluate blinks, it was used here a simple yet robust metric called eye aspect ratio (EAR), which does not require intensive signal and computer processing (such as EEG and EOG). Indeed, for the EAR evaluation, the proportion between height and width of the eye is calculated (i.e. higher values for open eyes and smaller for closed eyes) which easily extract a single feature from the whole image, reducing the amount of computational processing and storage memory required.

EAR has already been successfully adopted in works related with human alertness state (Soukupová and Cech 2016, W. O. Lee, Lee, and Park 2010). However, in this chapter, it was adapted and improved so that it could be applied not only for detecting blinks, but also to evaluate drowsiness in a real-time approach. Then, it would be possible to determine if a user (operator) is awake or sleepy based on the eye behavior.

To that end, the ML approach was here adopted. Although it is not fully clear which ML method performs better for activity recognition, SVM has confirmed successful application in several areas including heterogeneous types of recognition and pattern analysis (Anguita *et al.* 2012; Tripathy, Agrawal, and Rath 2016; I. D. Lins *et al.* 2013; Maior *et al.*, 2016). Moreover, compared with recent DL methods (GOODFELLOW *et al.*, 2016), SVM does not require powerful hardware and extensive training time, allowing usage on ordinary computers.

Thus, a useful tool for high-risk environments in which safety-critical tasks involve significant levels of alertness, such as industrial control rooms in process plants (e.g. nuclear and oil and gas) and airport traffic control towers is developed. Indeed, the majority of accidents (over 80%) in the chemical and petrochemical industries have human failure as a primary cause (BORGHINI *et al.*, 2014). Moreover, at the best of the author' knowledge, there is no work of this nature and approach applied to those fields. Furthermore, the model is able to run in standard computers and could be extended to almost every all-day work in which computers are required (e.g. financial services, database administrator, software developer).

3.2 COMPUTER VISION

CV is an interdisciplinary field aiming to investigate and develop software with high-level understanding from digital images or videos, describing the world that we see and to reconstruct its properties (SZELISKI, 2010). From the perspective of engineering, it seeks to automate tasks that the human visual system can do extracting information from images and videos (SONKA; HLAVAC; BOYLE, 2008). Information can mean anything from 3D models, camera position, object detection and recognition to group and search image content (JAN ERIK SOLEM, 2012).

CV gathers knowledge from many fields, such as image processing, pattern recognition, mathematics and AI. One of its main goal is to enable computers to reproduce core functions of human vision, such as motion perception and scene understanding. Taking advantage of these visual characteristics, computer vision is the feasible and appropriate technology to treat the problem of the drowsiness detection (FLORES; ARMINGOL; DE LA ESCALERA, 2010).

In this context, in the last couple of years, camera-based drowsiness detection devices using lid movement parameters have undergone intensive development. However, many interesting ideas do not seem to get beyond prototypes partly due to difficulties brought by large inter-individual differences (Jacobé de Naurois *et al.* 2018, Schleicher *et al.* 2008), which are characterized by distinct behavior when comparing different subjects. Distinct facial characteristics, either race or ethnics, could seriously impact model accuracy, and, thus, a feedback approach was here developed (see section 3.3.1.2.).

Fernandez *et al.* (2016) provide a valuable review of CV for system monitoring to detect attention. Authors define fundamental methods for face detection and tracking, as well as for facial landmark estimations, exploring algorithms for cognitive and visual distraction.

A specific challenge is to deal with the tracking problem. As stated by Zafeiriou, Zhang, and Zhang (2015), face detection is deeply studied not only because of the challenging nature of face, but also due to the countless applications that require face detection as a first step. The first practical, and yet well-known, algorithm to provide real time object detection was described by Viola & Jones (VIOLA; JONES, 2001). Even though several new tracking methods have been developed based on the cascade Viola-Jones algorithm (LEE; LEE; PARK, 2010), (WANG; GUO; CHEN, 2017), they still remain time consuming (JENSEN; LARSEN, 2008). One possible solution is to improve classical approaches for human detection such as scale invariant feature transform (SIFT) (JAIN *et al.*, 2015), local binary patterns (LBPs) (JUEFEI-XU; SAVVIDES, 2014) and histograms of oriented gradients (HOG) (BECERRA-RIERA; MORALES-GONZÁLEZ; MÉNDEZ-VÁZQUEZ, 2017).

Recently, deep convolutional neural networks (DCNN) appear as other possible solution, showing remarkable performance for object categorization (Krizhevsky, Sutskever, and Hinton 2012) with similar approach applied to face detection (VIJAYAN; SHERLY, 2019). DCNN and boosting based approaches using HOG/SIFT type features have common characteristics. Intuitively, it is possible to postulate that all DCNN layers contribute to the development of features invariant to face deformations as provided by classical methods (ZAFEIRIOU; ZHANG; ZHANG, 2015).

However, the gains in performance brought by DCNN have their price, which also limits its use. Indeed, DCNN requires a learning process from a large-capacity database and is computationally intensive using GPUs for fast parallel computation (Kim *et al.* 2017, Gu *et al.* 2018), which makes application in low-cost/standard computers difficult. In contrast, investigation of feature sets for human detection showed that locally normalized HOG descriptors provide excellent performance relative to other existing non-computational demanding feature sets (DALAL; TRIGGS, 2005) and is particularly suitable to characterize facial expression peculiarities (CARCAGNÌ *et al.*, 2015). As HOG descriptors are consistently used for object detection in CV, it was here adopted to perform facial landmark detection.

3.3 APPLICATION

Eyes, as one of the most salient facial features, reflect individuals' affective states, focus of attention and is considered a remarkable information source in face analysis (SONG *et al.*, 2013). Moreover, compared with yawns and eyebrow rising (JIMENEZ-PINTO; TORRES-TORRITI, 2012), eye-related metrics (e.g. blinks) are more reliable for drowsiness detection because gestures, such as laughing or talking with exaggerated lip movements, are easily misinterpreted, leading to false positives (BACIVAROV; IONITA; CORCORAN, 2008).

After performing face and eye detection, blink monitoring could be executed in order to track user alertness. An alternative for blink detection is using multiple cameras, which provides satisfactory results (ESPINOSA *et al.*, 2015). However, camera position alignment, calibration and apparatus demand large efforts to initialize models. Traditional blink detection methods using single and standard cameras (Salehian and Far 2015, John and Sharmila 2018) have also been proposed, but large distance from camera, illumination and high computational cost still seem to be a problem.

A common metric used to evaluate eye opening and closeness is the percentage of eyelid closure (PERCLOS) (MCKINLEY *et al.*, 2011). However, limitations in extracting pupil details, such as the problem of proper lighting, normally requires an infrared camera (MITTAL *et al.*, 2016). Moreover, PERCLOS pre-defines a threshold to characterize user attention (KANG, 2013), which could be insufficient considering the inter-individual characteristics.

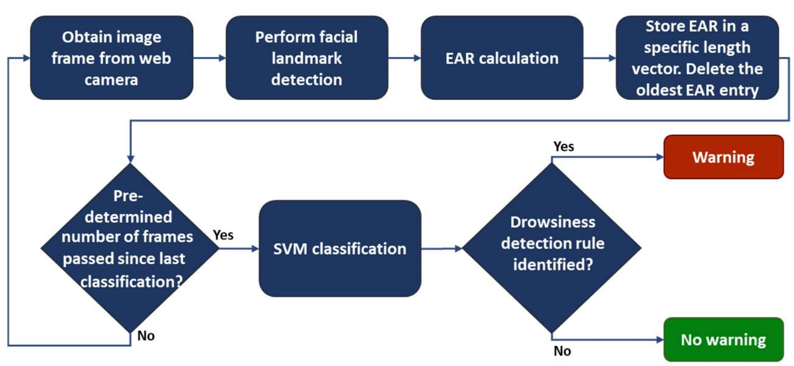
To deal with these problems, the use of simple, yet efficient, aspect ratio metrics (such as EAR) could represent an interesting alternative. For instance, Batista (2007) used face aspect ratio to estimate human face orientation and EAR to detect blink, emphasizing the variability of face and eyes shapes as an important challenging task. (LEE; LEE; PARK, 2010) and (RAKSHITA, 2018) also adopted EAR as one metric to detect blinks achieving interesting results related with robustness. However, authors used a fixed EAR threshold to determine a blink, which is not realistic dealing with the distinct people (i.e. inter-subject variety) and characteristics (e.g. natural eye openness). In Zhao *et al*. (2015), EAR is used to detect blinks, along with other eye-related variables. Even with satisfactory performance, authors complain about information processing velocity, which does not allow real-time usage.

In this chapter, low computational demanding methods to provide facial landmark recognition (i.e. HOG-based descriptors) and blink detection (i.e. EAR) was used, which allows real-time performance in standard computers. Moreover, none of the previously mentioned works relates blink pattern to drowsiness detection. Therefore, the developed approach allows usage of blink patterns analysis to warning user alertness in a real-time drowsiness detection.

3.3.1 Methodology

Figure 2 represents the core reasoning of the proposed methodology, which is structured on three main parts: (1) eye detection, (2) EAR calculation and blink classification; and (3) real-time drowsiness detection. Basically, EAR values are calculated and stored in each frame. To include the time dimension (i.e. not consider only a single image) and based on the duration of a standard blink (0.3-0.4 seconds, according to Bacivarov, Ionita, and Corcoran (2008)), a specific number of consecutive EARs were concatenated and used as input for an SVM model. Then, based on established blink pattern, it would be possible to determine whether the user is drowsy. If so, a sound and warning message should be emitted then. The first decision box in Figure 2 represents the real-time classification, i.e., the procedure of detection is repeated every pre-determined number of frames. The second decision box is related with the decision rule to detect and warn drowsiness. Further details for every main part are described next.

Figure 2 - Drowsiness detection methodology proposed for real-time monitoring

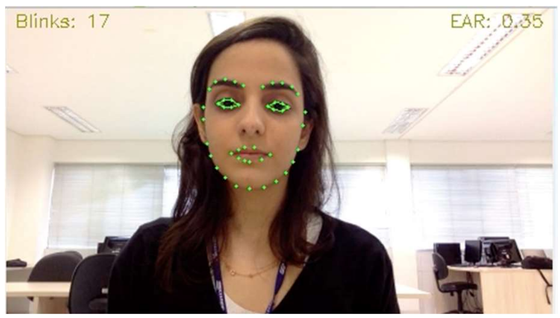


Source: The Author (2020)

3.3.1.1 Eye detection

The first step in constructing the real-time model for drowsiness classification is to locate the user's eye. To that end, dlib library (KING, 2003) was used for facial landmark estimation based on Kazemi and Sullivan (2014), which already adopts HOG and linear SVM. The library generates landmarks for the entire face, pictured as green dots in Figure 3. The landmarks are adaptable to different human faces and are not affected by considerable fast movements.

Figure 3 - Landmarks generated along the face.



Source: The Author (2020)

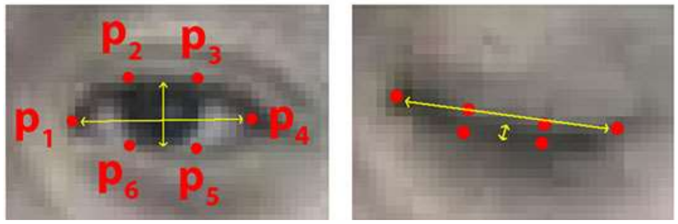
3.3.1.2 EAR calculation and blink classification

The previously described library provides landmark for the entire face, which are useful in other specific contexts. For example, information about head pose and gaze direction are used in literature to detect driver distraction (i.e. deviation from the road certainly represents a misbehavior (KAPLAN *et al.*, 2015). In the developed case (i.e. working with a computer), this specific information could mislead analysis because communication and interaction inside a control room, which requires deviation from the screen, should not necessarily be interpreted as misbehavior. Moreover, considering other face features may provide redundant information and increasing computational cost, which could compromises real-time monitoring with no consistent gain in drowsiness detection (JIMENEZ-PINTO; TORRES-TORRITI, 2012). Hence, in the model, only the eye-related landmarks are used.

Then, here, EAR, the proportion between width and height of the eye, was used based on its landmarks; see Equation (6). Figure 4 depicts all landmarks used in the EAR calculation. Unlikely to traditional CV methods for computing blinks, which normally implies in searching for the whites of the eyes and determine if they disappear for a period, it was possible here to easily (i.e. with low computational cost) extract an eye-based variable and make inference based on its value.

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{6}$$

Figure 4 - Eyes landmarks



Source: Adapted from Soukupová and Cech (2016)

EAR is estimated for each frame of the video streaming. Therefore, EAR decreases when user closes the eyes, increasing to a normal level when eyes are open again. This methodology is used to compute blinks as well as eye openness. For the construction of the drowsiness detection model, two different blink detection methods proposed by Soukupová and Cech (2016) were tested. However, modification and improvements were done for both methods to differentiate both short and long blinks.

The first method is called threshold method because a fixed limit was established and if user's EAR becomes lower than it for a particular number of frames (e.g. five consecutive frames for short blinks; 20 consecutive frames for long blinks), a blink is detected. Nevertheless, in practice, a single threshold does not suit every user since each person has a 'natural EAR' due to his (her) own face characteristics. A possible, and tested solution, was to set a calibration procedure to weigh EAR from a neutral face (e.g. when someone is reading) and a smiling face (e.g. when someone takes a picture). The aim was to capture EAR patterns from the natural face and from a distinct facial expression with subtle changes. Nonetheless, even after calibration, this approach leads to many false positive warnings because trivial expressions (e.g. talking) tended to decrease the EAR. Hence, this method was not considered robust and suitable for drowsiness detection.

Then, the second method has been considered: EARs from 15 consecutive frames were concatenated to create a user's state feature. This range of frames covers blink duration considering distinctive tasks possibly performed by users (e.g. reading texts, driving, performing vigilance task) as stated by Bacivarov, Ionita, and Corcoran (2008), Divjak and Bischof (2009), Ingre *et al.* (2006) and Kaida *et al.* (2006).

This feature, composed by 15 consecutive EARs, is classified into three categories: open eye, short blink or long blink (0, 1 and 2 respectively). Similarly to Maior *et al.* (2018), it was only considered a blink if the touch of eyelids occurred in the five central frames of the state feature (i.e. from the 6^{th} to the 10^{th} frame of the 15 selected frames). With this procedure, the same blink is prevented from being computed twice or more when the detection model is running in real time given that the model selects inputs every 5 frames.

Next, SVM is applied for classification purpose. SVM is a notable learning algorithm that creates an unknown mapping function of an input vector $\boldsymbol{x}$ and an class y based on the training data set D, not requiring any assumption or previous information about neither the function behavior nor the relation between input and output (WANG, 2005). Specifically, the model uses a 15-dimension input (15 consecutives EARs) and returns a category y (i.e. open eye, short blink or long blink).

Thus, an internal experiment was performed to generate the database used in the training process. Subjects were asked to read a text displayed on the screen during few seconds. Then, they were also demanded to stay with their eyes closed for a few seconds, representing the long blinks. The experiment aimed to identify both EAR behavior when subjects kept their eyes open and blinked (during the reading) as well as their closed eyes pattern related to long blinks.

From the total of 282 collected samples, 109 presented open eyes, 95 short blinks, and 78 were related to long blinks, comprising data of all possible state feature (i.e. 0, 1, 2). Yet, the database includes samples from distinct gender users with multiple physical characteristics (e.g. beard, mustache, glass wearing) also considering different head positions and facial expression. Information from each subject on age (avg. 34; median 30), gender (7 males; 6 females), beard/mustache (5 yes, 8 no) and use of glasses (5 yes, 8 no) are presented in Table 2.

Table 2 - Training database information

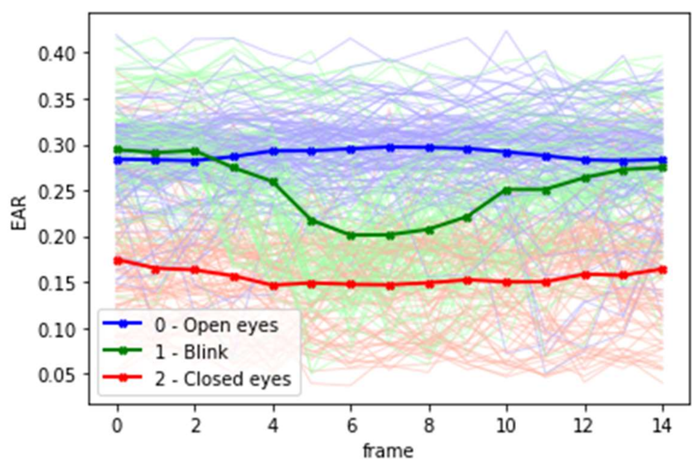
Subject	**Age**	**Gender**	**Beard or mustache**	**Glasses**
1	25	M	yes	no
2	27	F	no	no
3	53	M	no	yes
4	59	F	no	yes
5	23	M	yes	yes
6	22	F	no	no

Subject	Age	Gender	Beard or mustache	Glasses
7	23	M	yes	no
8	56	M	yes	no
9	21	M	no	no
10	37	F	no	no
11	33	F	no	no
12	33	F	no	yes
13	30	M	yes	yes

Source: The Author (2020)

Thirteen distinct users were used to provide EAR data since it usually differs from one person to another and, in that way, the algorithm was able to learn from varied situations (i.e. inter-individual differences). Figure 5 depicts every single input containing 15 EARs used in the training set, where solid lines represent averages in the three possible classifications. It is clear to analyze the expected eye behavior: for open eyes (blue), EAR is large and has little variation during the 15 frames; for short blinks (green) EAR quickly decreases during the closing of the eyes (which happens in the 5 central frames) and increases during the opening of the eyes; and for long blinks (red), EAR is small and has little variation during the 15 frames.

Figure 5 - EAR behavior for the three possible classifications.



Source: The Author (2020)

Next, SVM was trained with this dataset. Yet, SVM requires a set of parameters in its formulation and its adequate estimation also demands attention. Then, Optunity (CLAESEN; SIMM; JUMUTC, 2015), an optimization program, determined a suitable set of parameters that best adjusts the mapping function. Performing a grid search algorithm, SVM's parameters (C = 10; γ = 0.4) were estimated considering radial-basis function kernel. All methodology and algorithms here presented were developed in Python language, supported by the Scikit-learn package (PEDREGOSA; VAROQUAUX, 2011).

Then, a test dataset was created to evaluate the performance of the learned model when predicting unseen data. It is worth mentioning that all test data was generated in the same condition (i.e. same position of camera and small variations of proximity), although still considered different gender, glasses and beard/mustache, as well as illumination. The test dataset contains 98 samples, represented by 28 open eyes, 34 short blinks, and 36 long blinks.

Even though SVM model demonstrated consistent values in training and test phase (Table 3), an approach to further improve its efficiency was developed. The idea focused on the inter-individual differences, which may reduce model's performance due to variability of distinct users. Given that

short and long blinks decrease EAR, the challenge remains in defining how severe is this decrease to correctly indicate a blink. Indeed, EARs differences among distinct users mainly occur in open eyes, and hence this information is used to create a more personal model.

Specifically, during runtime, SVM pre-trained model automatically classifies user's new data (i.e. EARs sequences) into a particular class (i.e. 0, 1, 2). Then, EARs sequences classified as open eyes (i.e. 0) are stored until a set of a specific size (e.g. 20%) of training data is collected. It is important to emphasize that each EARs sequence classified as 0 is not necessarily selected, but only those surrounded by other sequences of 0 (i.e., data in possible intermediary states between open eyes and blinks is not used). That strategy was adopted to avoid that wrong classifications provided by the initial model highly influence the adjusted approach.

Then, this user-specific data is aggregated with existing training data to train a new SVM model. This process was called personal feedback because the initial SVM model is fed with individual information. Then, this updated model is used for state detection, now adapted for the current user. Considering the accuracy, the personal feedback model provides better predictions in all classes (i.e. 0, 1, 2), as presented in Table 3.

Table 3 - Accuracy of the proposed models

	Training (%)				**Test (%)**			
	0	1	2	Avg	0	1	2	Avg.

SVM	98.1	90.6	94.8	94.0	96.4	91.2	91.7	92.9
Feedback SVM	99.1	95.8	97.4	97.5	100.0	97.1	100.0	99.0

Source: The Author (2020)

3.3.1.3 Real-time drowsiness detection and warnings

It is important to mention that, even though classification uses an input of approximately 0.7 seconds (15 EARs) to give temporal notion, the model processes new sequences every 0.21 seconds (5 EARs) and one output of 'long blink' does not represent drowsiness. To infer about drowsiness state, more information is required.

Hence, to deal with drowsiness detection, it was not considered the classification of each output individually, but the whole pattern provided by them. Thus, SVM outputs (i.e. sequence of 0, 1 and 2) are used to determine whether the user is drowsy or not. Based on the related literature (STERN *et al.*, 1996), two different rules were considered as drowsiness:

Rule 1: If within a period of 60 seconds, the proportion between the '2' output (long blinks) and the sum of '1' and '2' outputs (total number of blinks) is higher than 25% (i.e. $\frac{\#2}{\#1+\#2} \geq 0.25$);

Rule 2: If 5 or more consecutive outputs (i.e. predictions of SVM) are 2 (long blink). This happens when the user stays with their eyes closed while face tracking is happening.

Fundamentally, the first situation occurs when the user slowly blinks several times in a short period. According to Caffier, Erdmann, and Ullsperger (2003), consecutive long blinks are highly correlated with first stages of drowsiness. The second rule in turn deals with continuously closed eyes (i.e., the user keeps the eyes closed for a long period). In this context, a lapse is defined as a failure to react or any reaction exceeding 500 msec and is often used as primary outcome measures of psychomotor vigilance tests (PVTs) (Belenky *et al.* 2003, Lim and Dinges 2008). Hence, for the second rule, the duration of two minimum lapses (i.e. at least one second) was considered.

If either situation is detected, a notification (visual and sound) pops up, warning the possible state of drowsiness, as presented in Figure 6. This message aims at warning users about their own state and avoid inattention, being refreshed in a user-defined minimum interval if the drowsy state remains.

Figure 6 - Drowsiness warning screen and notification.

Source: The Author (2020)

3.3.2 Model Validation

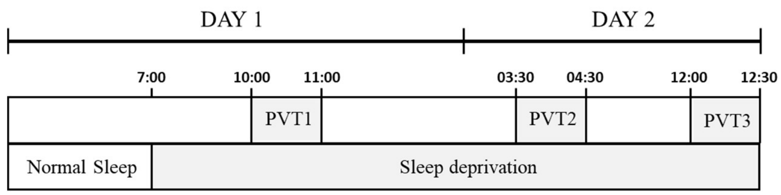
In order to validate the drowsiness model, tests were proceeded with a separated and public database. The "ULg multimodality drowsiness data-base", also called DROZY (MASSOZ *et al.*, 2016), is a database containing various types of drowsiness-related data (signals, videos, etc.) and intended to help researchers to carry out experiments, and to develop and evaluate systems (i.e. algorithms) in the area of drowsiness monitoring.

3.3.2.1 DROZY database

In DROZY, fourteen subjects independently performed three successive 10 minute long PVTs, during which a light was randomly turned on every few seconds and the subject should press a button as soon as the light appears, while their reaction time (RT) were registered (DINGES; POWELL, 1985). In safety-critical tasks, the PVT light may represent a situation when an alarm goes off in a control room showing a disturbance in the operation process (e.g. temperature and/or pressure rise).

These PVTs occurred in two consecutive days under conditions of increased sleep deprivation induced by acute and prolonged waking (see Figure 7). In fact, the total sleep deprivation time between the first and third PVTs was about 28 to 30 hours (MASSOZ *et al.*, 2016).

Figure 7 - PVT procedure in DROZY.



Source: The Author (2020)

In DROZY, subjects were also asked to fulfill a Karolinska sleepiness scale (KSS) (ÅKERSTEDT; GILLBERG, 1990) form, which is a widely established method to measure the subjective level of sleepiness at a particular period of the day (KAIDA *et al.*, 2006). Indeed, KSS is a validated assay of subjective sleepiness, highly sensitive and commonly used for assessing overall tiredness (Åkerstedt *et al.* 2014, Ding *et al.* 2019, Wilson *et al.* 2019). The KSS scale is presented in Table 4 (ÅKERSTEDT *et al.*, 2014). Thus, subjects indicate which level best represents their psychophysical state at the beginning of PVT.

Table 4 - KSS Scale for auto evaluation of drowsiness

Karolinska Sleepiness Scale (KSS)	
1	Extremely Alert
2	Very Alert
3	Alert
4	Rather Alert
5	Neither Alert nor Sleepy
6	Some Sights of Sleepiness
7	Sleepy, But No Effort to Keep Awake
8	Sleepy, Some Effort to Keep Awake
9	Very Sleepy, Fighting Sleep

Source: The Author (2020)

The conventional mental fatigue measurement methods can be classified into two categories: subjective and objective. In the former, subjects rate their level of mental fatigue (e.g. KSS), while the latter assesses mental fatigue via quantifying subjects' performance for a specific task (e.g. RT in the PVTs) (Sandberg *et al*. 2011, Cole 1982, Lefkovits, Lefkovits, and Emerich 2017). There is a general agreement that these conventional measurement methods can have good validity and reliability (ZHANG *et al.*, 2017). The usage of RT as well as the KSS as measurement of alertness performance is well-known in literature (FRANÇOIS *et al.*, 2016), which allows the use of both information provided by DROZY.

In the validation procedure, the model uses videos as real-time input to detect whether the subject is drowsy or not; Figure 8 exemplifies one of the videos used. Even though initially containing 14 subjects, each one performing three PVTs (PVT1, PVT2, PVT3), DROZY does not provide complete information (e.g. videos; PVTs performance) of all experiments, which reduces the amount of useful data in those cases.

Figure 8 - Validation video from DROZY

Source: The Author (2020)

Yet, it is worth mentioning that in DROZY, the camera was located just below the screen, a different position from a standard computer cam (above the screen), in which the model was trained. This situation caused the eye tracking to be very sensitive. Moreover, videos were recorded in different frames per second (15 and 30 FPS), which is a few frames different from the one used (~ 23 FPS) to train the model, turning out to be an even more challenging validation set. Despite those remarkable differences, the model could still provide interesting results, as shown next. For further information about the experiment and DROZY dataset, see Massoz *et al.* (2016).

3.3.2.2 Inter-subject results

According to Table 4, it is expected that lower KSS levels correspond to situations where the subjects do not have reduced performance due to drowsiness, while higher KSS levels represent possible states of sleepiness. Then, the model evaluates people in different stages of alertness (e.g. awake and drowsy) such as in Samiee *et al.* (2014). In this way, there were 8 subjects who classified themselves as one of the first three stages of alertness (levels 1, 2 or 3 in Table 4) when performing PVT1. DROZY also contained 10 videos in which subjects believe to employ some or huge effort to stay alert, representing one of the last three phases of KSS (levels 7, 8 and 9) related with videos during PVT3. The specific KSS classification for each object could be seen in Table 5.

To tackle the FPS variability, adaptation of the model to process videos with its specific recorded rate (e.g. 15 and 30 FPS) were done, with all analysis lasting around 10 minutes. This approach represents the 'real duration' of the experiment, independently of the FPS rate of the camera.

To evaluate the aggregated performance (i.e. considering inter-subject analysis), PVT1, performed at the beginning of the drowsiness experiment, and PVT3, carried out at the end of the same experiment, were investigated for each subject. The ground truth of no warning for subjects in PVT1 and at least one warning for subjects in PVT3 is defined for each subject. Given that, the model warned correctly in 94,44% of the videos, which demonstrates robustness despite the FPS change. The number of warnings for each PVT video in DROZY database are also shown in Table 5, which provides an important analysis to verify the model consistency.

Table 5 - KSS classification and number of warnings emitted for all subjects and PVTs considered

	PVT1		PVT3	
SUBJECT	**KSS**	**warnings emitted**	**KSS**	**warnings emitted**
1	3	0	7	11
2	3	0	-	-
3	2	0	-	-
4	-	-	9	2
5	3	0	8	0
6	2	0	7	19
7	-	-	9	33
8	2	0	8	1

9	**-**	**-**	**8**	**26**
10	**3**	**0**	**7**	**7**
11	**-**	**-**	**7**	**28**
12	**2**	**0**	**-**	**-**
13	**-**	**-**	**-**	**-**
14	**-**	**-**	**8**	**34**
MEAN	**2.5**	**0**	**7.7**	**16.1**

'-' denotes a video not analyzed (i.e. KSS was not 1, 2 or 3 for PVT1 or 7, 8 or 9 for PVT3).

Source: The Author (2020)

Despite tracking difficulty due to different camera position and image characteristics - color, brightness, presence of variety of electrodes (e.g. compare Figure 6 and Figure 8) - no warning was emitted for any subject performing PVT1, with zero false positive. Moreover, in none case PVT1 had more warnings than PVT3, which is compatible with the increase of drowsiness. Regarding sleepiness, in 90% of the PVTs, the model warns signs of drowsiness. Moreover, looking at the mean number of warnings for PVT1 and PVT3, the difference is considerably high, which is in accordance with the KSS and the experiment described in DROZY.

3.3.2.3 Specific alert-drowsiness subject results

Further analysis was done for specific subjects that went from alertness (levels 1, 2, 3) to drowsiness (levels 7, 8, 9) states through PVTs, which allows a direct comparison between their performance in both states. In other words, four subjects (i.e. subjects '1', '6', '8', '10') for which the drowsiness level increase from PVT1 to PVT3 were analyzed. Their performance on each PVT and the model's warning are summarized in Table 6.

Table 6 - Subjective and objective drowsiness-related metrics for all specific subjects analyzed

Subject	PVT	Subjective metric	Objective metric			# warnings emitted by the model
		KSS Level	*# Lapses*	*RT mean (in s)*	*RT variance (in 10^{-3} s)*	
1	**PVT1**	**3**	**0**	**0.25**	**2.10**	**0**
	PVT3	**7**	**1**	**0.29**	**8.34**	**11**
6	**PVT1**	**2**	**2**	**0.34**	**1.94**	**0**
	PVT3	**7**	**40**	**0.50**	**12.63**	**19**
8	**PVT1**	**2**	**3**	**0.35**	**2.44**	**0**
	PVT3	**8**	**23**	**0.45**	**12.79**	**1**
10	**PVT1**	**3**	**3**	**0.35**	**2.25**	**0**
	PVT3	**7**	**7**	**0.39**	**6.92**	**7**

Source: The Author (2020)

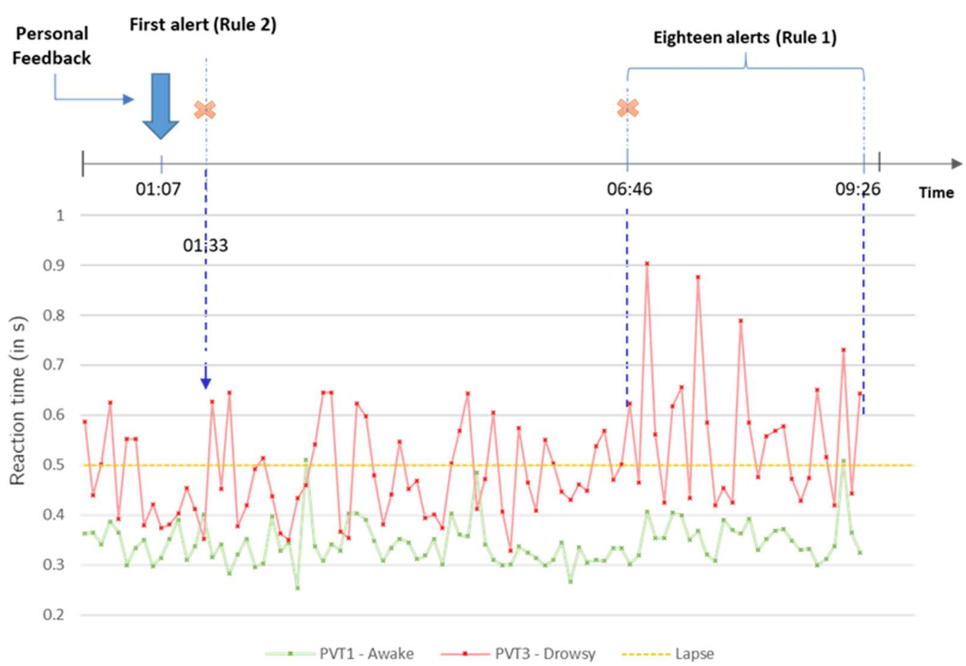
As previously mentioned, subjective and objective information to confirm the drowsiness state of a subject were gathered. As a subjective (or qualitative) metric, KSS evaluation was used, while for objective (or quantitative) metric, RT mean, RT variance and the number of lapses (i.e. when RT is greater than 0.5 seconds) were analyzed. Moreover, the Wilcoxon Signed-Rank test for paired samples (HOLLANDER; WOLFE; CHICKEN, 2014) verified whether the median of the differences between RTs in PVT3 and PVT1 was greater than 0 for each of the four subjects. Considering a significance level of $\alpha = 0.05$, the null hypothesis was rejected in all four cases (p-value < 0.05) confirming that RTs are greater when the subject is drowsy. Thus, these metrics were used as performance indicators as it is commonly seen in drowsiness experiments (Peiris *et al.* 2005, Kaida *et al.* 2006).

For example, in PVT1 of subject '6' (KSS = 2), the RT mean was 0.34 seconds, while this average increases more than 47% up to 0.5 seconds in PVT3 (KSS = 7). Further, in PVT1, only in 2 instances the user had RT higher than 0.5 seconds (i.e. lapse), while in PVT3, this value rises 20 times, up to 40, which confirms the decreasing performance when drowsy. For this subject, the model provided drowsiness warnings 19 times in PVT3, whereas no alert was emitted during PVT1, being very consistent with the subject state.

The performance of the remaining three subjects could also be seen in Table 6. Here, comments on the intrinsic inter-subject variance, which normally provides individual results, can be done. For example, subject 1 presented just one lapse during PVT3, even though its KSS was 7. Also, its mean RT only increased 0.04 seconds between PVTs, which implies that even feeling drowsy, performance remained quite effective. Despite that, the model presented warnings for PVT3 and none was emitted for PVT1, which is considerably robust. A similar analysis could be done for the other subjects in Table 6.

Other interesting characteristics are observed looking at the variance of the RTs. For subject ‘6’, for example, an increase of more than 6 times is seen between PVT1 and PVT3 (i.e. from 1.94 to 12.63 - in 10^{-3} s), which is clearly depicted in Table 6. This behavior indicates that, during PVT3 (almost 10 minutes), subject ‘6’ greatly varies its performance, even though it is generally worse as compared with PVT1. Moreover, during PVT3, there are moments when subject ‘6’ is even more prone to drowsiness. Figure 9 helps to understand this pattern and how the model identified which are the critical moments of drowsiness.

Figure 9 - Timeline of warnings emitted by the model (superior) and the RT (inferior) of subject ‘6’ (green line denotes PVT1; red line denotes PVT3).



Source: The Author (2020)

Firstly, difference in performance is easily observed when comparing green and red lines (i.e. RT in PVT1 and PVT3, respectively) in Figure 9. Moreover, looking only at the red line, one can identify a worse behavior happening in the last part of the test. For this particular PVT, during the interval of 06:46 to 09:26 (i.e. 3 minutes), subject ‘6’ presented 17 lapses (i.e. mean of 5.66 lapses/minute), which is more than 40% higher than the average number of lapses per minute in the video (4 lapses/minute). This represents a specific and critical stage, where the subject had greatly lost performance. During the same interval, the model warned drowsiness 18 times due to Rule 1, which indicates a continuous stage of drowsiness. One alert was emitted once in the beginning of video due to Rule 2, which could possibly be seen as an early pattern of drowsiness. A similar analysis could be done for the other subjects, with the model remaining consistent with the KSS evaluation and performance provided by the RT.

3.4 DISCUSSION

An ML-based model to automatically identify drowsiness patterns of a user is here proposed. The method uses face landmarks to estimate EAR, and then applies SVM to classify the state feature in a low-cost processing real-time approach. This SVM model was upgraded with specific information from the user by means of personal feedback. Then, decision rules derived from studies of neuroscience were adopted to determine whether the user was drowsy or not.

The model was externally validated using the database DROZY. In inter-subject analysis, the model provided 94,44% of accuracy, considering whether subject was awake or drowsy. Moreover, for specific subjects, the approach strongly alerts about crucial moments in drowsiness state, proving to be efficient and reliable using solely an ordinary web camera. Still, the application is physically non-intrusive and could be run alongside other programs in a personal computer without noticeable practical performance impact, allowing to enhance safety in monotonous activities that demand high levels of alertness.

Despite the difference in FPS and position of camera, the model presented great results on DROZY. However, there is still room for improvement. Indeed, given that the model was developed to be used for a standard computer web camera (often located at the top of the screen), track missing (e.g. no alerts for subject ‘5’) was noticed in the validation test. Specifically, with this kind of view (bottom), tracking was not as robust as it is from the top view, which makes the model present poorer performance when compared with the internal validation done in the training conditions. It is expected to overcome problems by enhancing tracking algorithm of bottom parts of the face (e.g. chin, neck), without losing real-time performance.

Furthermore, it is important to mention that even if subjects classify themselves with some sort of sleepiness (i.e. stages 7, 8 or 9 in KSS) and present high RT, this does not necessarily mean their eyes’ pattern presented image evidence to reveal their actual alertness state. People often try to compensate effects of fatigue so that performance appears to be normal (RAMOS *et al.*, 2017) (e.g. keeping the eyes widely open). This is a limitation of the model since it uses only images (videos) as drowsiness evidence.

As a future work, comparison from monitoring of other facial landmarks (e.g. mouth tracking) in the drowsiness detection method (OMIDYEGANEH *et al.*, 2016), still keeping real-time processing should be done. Yet, a possible consideration is to train the model with image/videos with camera in different positions (e.g. below as in DROZY) and different FPS. Moreover, another approach is to construct the model with longer state features of EARs (e.g. using 25 frames instead of 15) in order to provide even more robustness about the actual drowsy state, being careful to notice that it possibly would enhance the computational cost.

4 METHODOLOGIES FOR ESTIMATION OF REMAINING USEFUL LIFE FROM VIBRATION SIGNAL

Some results discussed on this chapter have been published on the journal Eksploatacja i Niezawodność - Maintenance and Reliability and on the European Safety and Reliability Conference - Hannover in 2019.

4.1 CONTEXT

Nowadays, growth on the expected demand of distinct industrial sectors often requires uninterrupted production, keeping the operating system usage as higher as possible. In those complex systems, a fault in a specific component quickly produces chain reaction and induces damage on other components. Such unexpected faults lead to machine break down, resulting in economic loss and even person safety threat (GUO *et al.*, 2017).

Hence, technology had rapidly advanced to create a variety of cost-efficiently sensors and tools to provide real-time information (i.e. data) of specific systems (MAIOR; MOURA; LINS, 2019). Specifically, the massive amount of industrial data can be used for monitoring purposes, as well as to determine the health state of a system and thus support implementation of preventive actions before catastrophic failures, providing new possibilities for improvement of reliability and efficiency (Cofre-Martel *et al*. 2019, Ren *et al*. 2017).

In this context, Prognostic and Health Management (PHM) aims to monitor degradation in engineering systems, understand when failure may occur and provide a cost-effective strategy for scheduled maintenance (MAIOR; MOURA; LINS, 2019). One of the main challenges faced by industries is early detection of faults in machinery (SAN MARTIN *et al.*, 2019).

System prognosis is a key factor within the condition-based maintenance strategy and has been highlighted in different fields of science (Widodo and Yang 2011; Sutharssan *et al.* 2012). In this context, Remaining Useful Life (RUL) is a rather common measure used to characterize equipment performance (SIKORSKA; HODKIEWICZ; MA, 2011). According to Si *et al.* (2013), RUL is the useful life left at a particular time of operation, and is typically random and unknown. In fact, RUL is related to several factors (e.g. current degradation state, operating environment, system function) and should be estimated from available sources of information such as condition and health monitoring sensors.

Different signals can be collected in order to track the degradation of a system, and then build an accurate relationship between the current health condition state and RUL. Many signals (e.g. vibration, acoustic emission, temperature) can represent the evolution of degradation, and their analyses are as necessary as arduous (Chang *et al.*, 2018; El-Thalji and Jantunen, 2015; Ambhore *et al.*, 2015).

However, the enormous number of environmental, process and operating variables influencing the desired system (e.g. health status of a machinery) makes the determination of specific contribution of each one impracticable. Moreover, more complex, high-dimensional and noisy real-world time-series data cannot be described with analytical equations with parameters to solve since the dynamics are either too complex or unknown (TAYLOR *et al.*, 2010). Even though there is no universally accepted best model to estimate RUL (LIAO; KÖTTIG, 2014), current promising statistical methods have dealt with real-time big data (BOUSDEKIS *et al.*, 2015).

4.2 APPLICATION

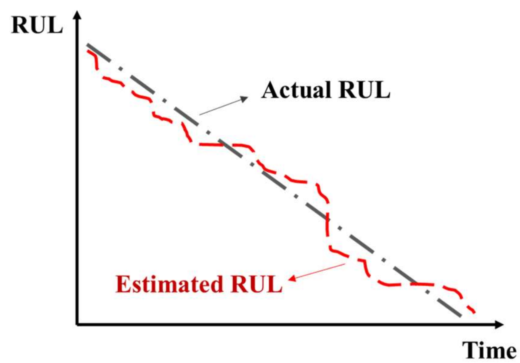
In this chapter, two approaches were considered: an ML-based methodology and a DL-based methodology. Both approaches were applied to a real bigdata set provided by FEMTO-ST Institute generated in the IEEE PHM 2012 Data Challenge focused on the estimation of the RUL for bearings based on vibration signals (NECTOUX *et al.*, 2012). The data results from experiments carried out on a laboratory experimental platform (PRONOSTIA), that enables accelerated degradation of bearings under constant and/or variable operating conditions, while gathering online health monitoring data (e.g. vibration).

The experiment main objective was to provide data that characterize the degradation of bearings along their complete operational life (until their total failure). Yet, considering the nature of a PHM challenge, data was complex and tricky, which certainly jeopardize the prediction capacity of the proposed models. The database have become popular during recent years in distinct types of applications (Ren *et al*. 2018; Mao *et al.* 2018, Fumeo,Oneto, and Anguita 2015; Boškoski *et al*. 2015). For further information about the dataset, see Nectoux *et al.* (2012).

This application aims to replicate exactly the IEEE PHM 2012 Data Challenge, which presents unclear end-of-life signature and unbalanced dataset (HUANG *et al.*, 2017), for two specific bearing. In this case, one bearing, here called 'Train Bearing', was used for training purpose and presented vibration data until its failure. It had 2,803 recordings, each one containing 2,560 points for horizontal vibration and other 2,560 points for vertical vibration. Vibration data provided for other bearing, here called 'Test Bearing', consisted of 1,802 records, also with 2,560 horizontal and vertical vibration points, each. In this case, only truncated data was provided and the challenge was to estimate the correct RUL based on its initial degradation behavior.

For both ML and DL methodology, it is not expected the direct point prediction to be enough precise due to the high variability of the data. However, the trend of all predictions should express the realistic RUL estimation, as seen in Figure 10 (SUTRISNO *et al.*, 2012). Comparison of ML and DL models are presented in section 4.2.3.

Figure 10 - Expected estimated RUL behavior.



Source: Adapted from Sutrisno *et al.* (2012)

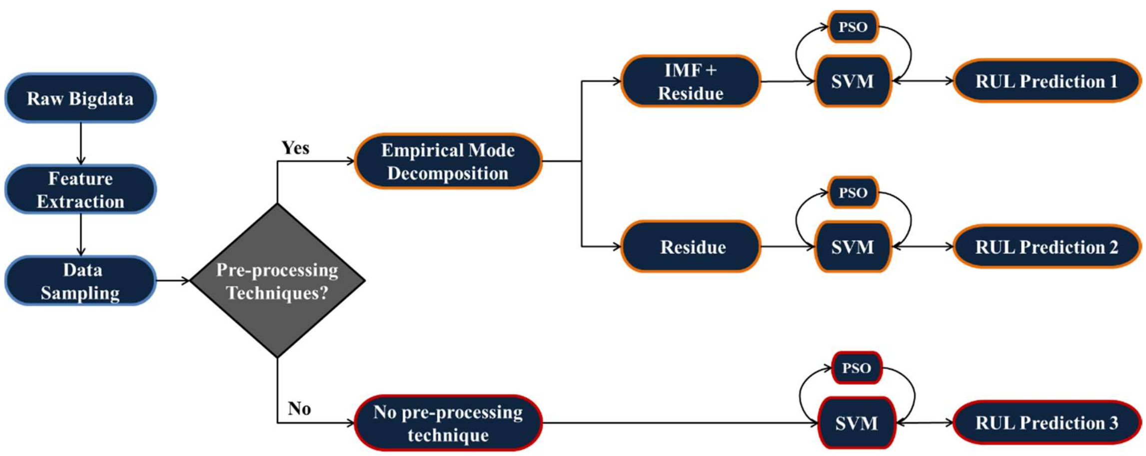
4.2.1 ML methodology

As in this thesis, several SVM-based models (Soualhi, Medjaher and Zerhoun 2015; Saha, Goebel and Christophersen 2009; Patil *et al.,*2015) have been proposed to predict RUL. However, once SVM learning performance strongly depends on the quality of the input data and the direct use of the original series as input variables could consider irrelevant information (e.g. noise) and/or miss important features, pre-processing methods could be used to improve data input quality (MAIOR *et al.*, 2016).

A notable pre-processing technique is Empirical Mode Decomposition (EMD), which decomposes the original series into a sum of Intrinsic Mode Functions (IMFs) and a final residue. Each IMF represents a frequency-amplitude modulated narrow band, normally associated with a specific physical process, while the residue has intrinsic attributes related with flatness and physical meaning of the signal trend (YANG; YU; CHENG, 2007). According to Huang *et al.* (2014), EMD is adaptive, empirical, direct and intuitive. For further information about EMD and others pre-processing techniques applied in PHM context, see Maior, Moura, and Lins (2019).

The proposed methodology for the ML approach is presented in Figure 11. Here, EMD-preprocessed models where analyzed when coupled with optimized-SVM. Here, Particle Swarm Optimization (PSO), a probabilistic optimization heuristic inspired in the social behavior of biological organisms, is used to provide the best set of parameters to be used in SVM. PSO-optimized SVM have been successfully applied in reliability problems (Droguett *et al.,* 2014; Lins, Moura and Droguett, 2013; García Nieto *et al.,* 2015). A model with no EMD preprocessing is also evaluated for comparison purposes.

Figure 11 - Methodology for the ML approach



Source: The Author (2020)

Considering the large quantity of information and its inherent computational cost, this learning model cannot directly handle such an extensive data. Hence, to cover a massive data in ML methodology, two previous steps were performed for data dimension reduction: (i) feature extraction and (ii) data sampling. The former aims to reduce a series of points into a representative measure (e.g. mean, kurtosis, the highest absolute value), while the latter consists in sampling from original data. Indeed, for healthier states of bearing, lower sampling frequency is necessary, while for more degraded states, higher sampling frequency is required.

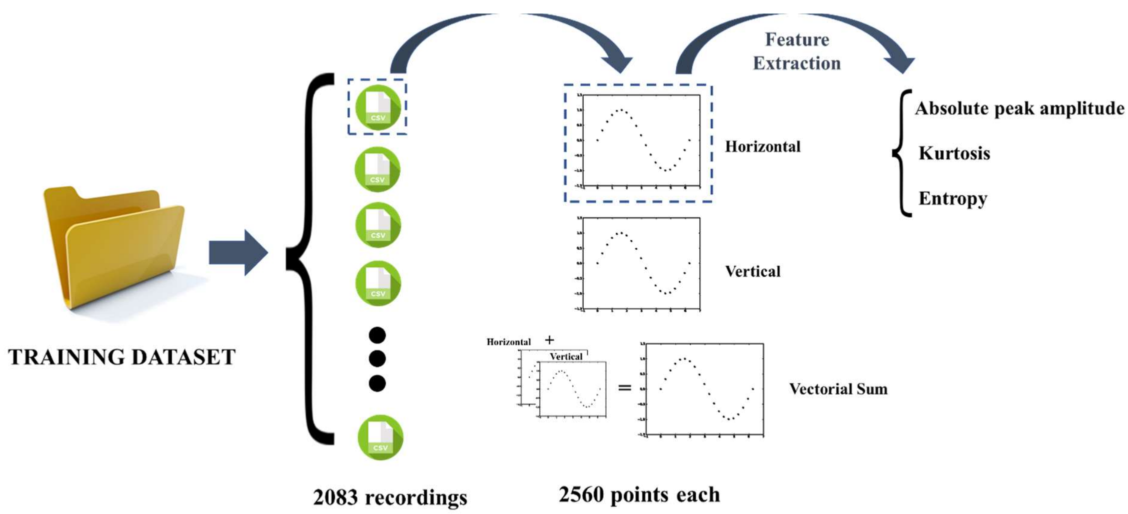
After sampling, EMD is performed. Here, two distinct regression models were created: one model contained all IMFs and the residue, while the other model contained just the final residue.

The next step was to input PSO+SVM model with the previous processed data. Then, evaluation of each model (1 – IMFs + Residue; 2 – Residue; 3 – No pre-processing) based on the performance of RUL prediction was done.

SVM supervised learning method requires both y (i.e. the response variable) and $\boldsymbol{x}$ (i.e. the regressor/input) variables. In all ML cases, the response variable was the RUL and the regression variables were the vibrations signal related with a specific bearing.

Considering the 2083 recording (horizontal and vertical), more than 14 million points are provided only for training purposes. Also, in this methodology, a third signal composed by the vectorial sum of the horizontal vibration and vertical vibration was analyzed. For each of the three vibration signals (i.e. vertical, horizontal and vectorial sum), three metrics were calculated in the feature extraction step: absolute peak amplitude, kurtosis and entropy. Figure 12 summarizes the data provided and the feature extraction process for the large amount of information.

Figure 12 - Feature extraction process considering the dataset.

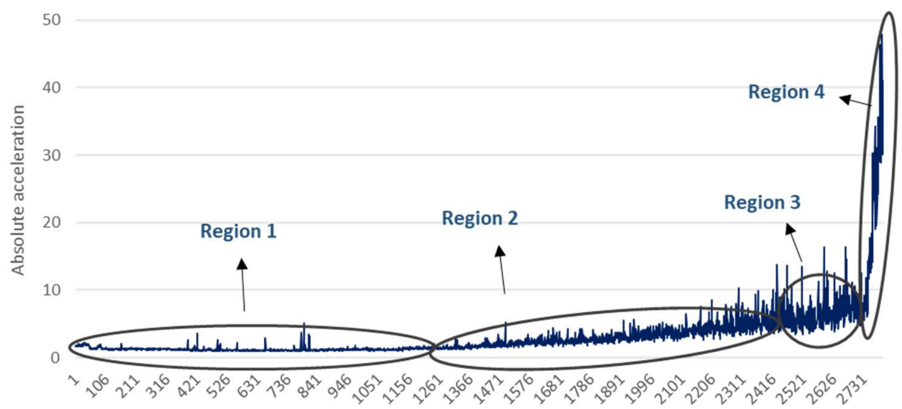


Source: The Author (2020)

Here, investigation about function variation over time as well as the expected behavior when bearing degrades indicated the absolute peak amplitude as the most suitable feature to be used in the next steps, similarly to previously published works (e.g. Rohlmann *et al.*, 2014; Chen *et al.*, 2017). Specifically, when considering just a small amount of data (i.e. 10% of total) the horizontal vibration signal presented better initial results and was the chosen one to be analyzed. Moreover, for the absolute amplitude, the average of the five highest absolute peak acceleration values in each recording was considered. Averaging was done in order to alleviate the effect of data noise (LEE; YUN, 2006).

After feature extraction, the data were categorized in four different regions in similar procedure to ISO 10816 (STANDARDIZATION, 2009) that deals with condition monitoring based on vibration. Therefore, each region represents a degradation phase of the bearing, in which a more degraded phase presents a higher vibration amplitude. A change of region takes place when the vibration trend line for the current region suffers a sudden increase of inclination (e.g. a new crack appears). The four regions are shown in Figure 13.

Figure 13 - The four different regions of degradation.



Source: The Author (2020)

In order to reduce the amount of information, data sampling was performed in every region with distinct sampling frequency (i.e. the more unstable the bearing is, the more necessary is to monitoring). Table 7 depicts the sampling frequency, the total duration for each degradation region and the number of points actually used after sampling. To illustrate, the third region was sampled every 100 seconds during the 2500 seconds in which the bearing stayed in this region, providing a total of 25 points. Hence, a total of 137 points were used for training purpose.

Table 7 - Sampling frequency and duration for each degradation region

Degradation Region	**Sampling Frequency (in seconds)**	**Total time (in seconds)**	**Number of points considered**
1	400	12000	30
2	200	13000	65
3	100	2500	25
4	30	510	17

Source: The Author (2020)

'Test Bearing' presented 1803 recordings and it is expected the bearing to pass through all four degradations regions, even if the truncated data of does not present all of them. Indeed, 'Test Bearing' available data does not present several abrupt changes in the signal, seemingly representing only first and second degradation regions. Therefore, inference had to be done to further degradation zones.

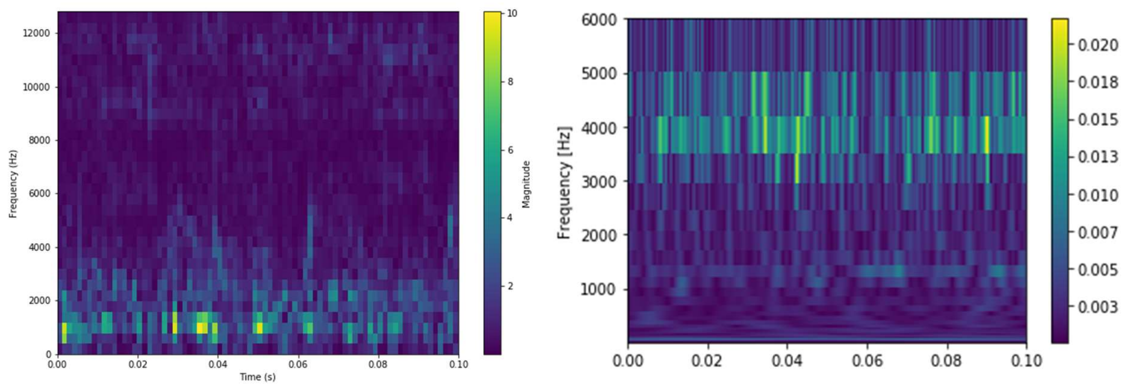
As depicted in Figure 13, vibration from the healthier stage (i.e. first degradation region) is almost stationary, with negligible fluctuations, even though it represents a considerable amount of life data. For 'Test Bearing', which presented truncated, this region represents even a higher proportion of the total available data. Given that, using data from region 1 will not represent any gain about bearing degradation. Moreover, using data from region 1 will only deviate the overall trend.

Therefore, in the test case, no data from the first region was used. Also, sampling rate of the second region was adapted to provide more information, i.e. frequency of the fourth region was used (every 30 seconds). Thus, test data was reduced to just over 100 (i.e.120) test points in which RUL would be predicted based on its vibration signal. As previously mentioned, the final RUL estimation is based on the overall trend of the predictions. Results of this ML methodology can be seen in Maior, Moura, and Lins (2019) and is presented in section 4.2.3, along with comparisons with DL models of section 4.2.2.

4.2.2 DL methodology

As an alternative approach, in this DL methodology, CNN models were created to predict RUL of the same bearings of section 4.2.1. As previously mentioned, CNN has a remarkable performance working on images. Once the raw data is provided in the form of a time series vibration signal, it is convenient to adequate the input by converting it in a 2-D representation. For that goal, two techniques based on time-frequency transformation were applied and compared: Short-Time Fourier Transform (STFT) and Wavelets Transform (WT). Figure 14 depicts an image input representation to be used, where the x- and y-axis are time and frequency, respectively, and the color scale indicates the frequency amplitude.

Figure 14 - Image input data provided by (a) STFT spectrogram and (b) WT scalogram.



(a) (b)

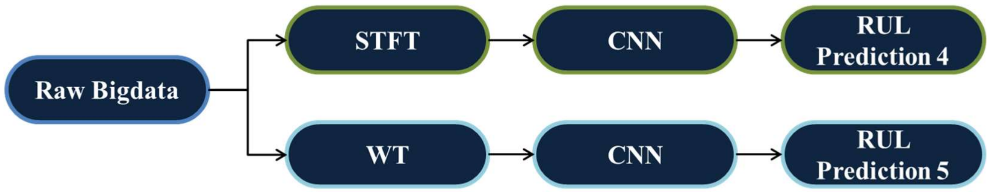
Source: The Author (2020)

The basic idea of STFT is to divide the initial signal into small time windows and apply the Fourier transform to each time segment for representing the variation in signal frequency content over time that existed in that segment (GOYAL; PABLA, 2016). STFT has a fixed resolution independent of the width of window selected, which demands a trade-off between a poor time/good frequency resolution or a good time/poor frequency resolution. The basis for the STFT representation is a series of sinusoids and it is common and direct frequency domain analysis. For a review of STFT, and its variation, applied vibration signals, see Lin and Ye (2019).

Wavelet Transforms (WT) was first proposed by Morlet *et al*. (1982), introducing the idea of fine grained time–frequency analysis approach to achieve an optimal balance between frequency resolution and time resolution (GOYAL; PABLA, 2016). WT converts a signal of time domain using a wavelet basis function into a different form (Mallat 2009; Yan, Gao, and Chen 2014), becoming a linear time-frequency representation with a wavelet basis instead of sinusoidal functions. As wavelet basis, this thesis applied the Morlet wave once it has frequently and efficiently been used for time-frequency analysis of non-stationary time series and transient signals (COHEN, 2019). For a review of WT, and its variation, see Chen *et al.* (2016).

Adjusts on the ML methodology (Figure 11) were made for the DL methodology, consisting in a more straightforward form (Figure 15). In a similar way to preprocessing techniques in ML approach, STFT was used to transform the vibration data to a suitable form for DL approach. However, here, other pre-processing steps (i.e. feature extraction and data sampling) were not performed.

Figure 15 - Methodology for the DL approach

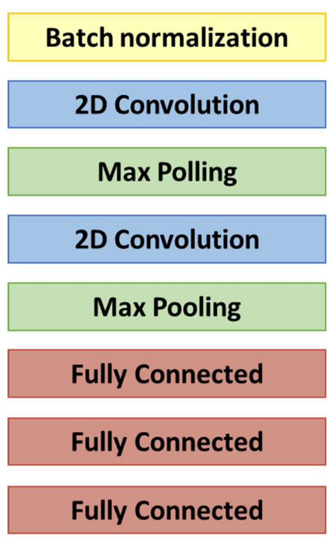


Source: The Author (2020)

Note that the elimination of human-performed feature extraction and data sampling phases requires a lot less effort to generate the models. Unlike the ML methodology, here the inputs are not vibration data but the time-frequency images generated by STFT and WT. The same bearings of section 4.2.1 was used for training and test phases. In this case, each 2083 recording of ‘Train Bearing’ was applied to STFT and WT, generating 2083 images for each case. Similarly, ‘Test Bearing’ had 1803 recordings, which creates 1803 images for test predictions. Once again, the bearing RUL estimation is derived from the trend provided by the 1803 test predictions.

In this DL approach, the architecture of CNN highly impacts the performance. For both STFT and WT applications, an elementary architecture was applied composed of convolution, pooling and fully connected layers, as depicted in Figure 16. An initial batch normalization layer is also used to normalize the initial data and network weights, and to avoid overfitting.

Figure 16 - CNN architecture for the DL methodology.



Source: The Author (2020)

Specifically, in the first and second convolutional layers, 8 and 4 feature maps were used, respectively. For the fully connected layers, 8, 4 and 1 neurons were used, respectively. For all convolutional and fully connected layers, 'relu' activation function was used due to its improvements in DL models (FRED AGARAP, 2018). Both models were trained during 1,000 epochs in a i9 9900 processor, 32 GB of RAM computer with a Nvidia GEFORCE GTX 2080 Ti GPU. STFT training model took approximately 20 minutes, while WT took approximately 90 minutes.

4.2.3 RUL prediction results

In order to compare models, Absolute Percentage Errors (APE) were calculated, quantifying the distance error from the real RUL to the estimated one. To define RUL estimation for the challenge, the trend line of the predictions (e.g. Figure 10) was extrapolated until it crosses the x-axis (i.e. trend estimates RUL equals to 0). Table 8 presents the models' performance of ML and DL models for which the correct RUL was 5730 seconds.

Table 8 - Errors for all tested models

	Model	**Regressors**	**APE**
ML Models	1	IMFs + Residue	15.39%
	2	Residue	24.90%
	3	Direct Vibration Data	58.53%
DL Models	4	STFT+CNN	29.32%
	5	WT+CNN	14.92%

Source: The Author (2020)

For ML models, EMD preprocessed models (Models 1 and 2) predicted best results when compared with the model with no preprocessing one (Model 3). Here, it is important to emphasize the difficulty of this challenge, which is evident by considering the magnitude of errors in Table 8. Indeed, the winner of the IEEE Challenge, based on the same evaluation metric, presented prediction errors of 37% (SUTRISNO *et al.*, 2012). The winner's prediction is worse than estimations provided by two of the ML presented models. Moreover, the Model 1 reduced the error in more than 58%, which confirms the advantage in using the pre-processing method proposed. These results are published in Maior, Moura, and Lins (2019).

Table 8 also brings interesting results for DL models. STFT and WT-based model presented results comparable to pre-processed ML models without the necessity of manual and burdensome steps (i.e. feature extraction and sampling). Moreover, WT model presented the best performance overall, which is an achievement for this challenge. However, it is important to notice that DL calculations were only possible to be done in a reasonable time due to a powerful GPU usage and considerable RAM storage, which, today, could limit the similar applications.

For future research, investigations about others preprocessing techniques should be analyzed to improve ML performance. Moreover, comparison with variants of EMD techniques (e.g. Ensemble Empirical Mode Decomposition (EEMD) (WU; HUANG, 2009), Complete Ensemble Empirical Mode Decomposition (CEEMD) (TORRES *et al.*, 2011) could be done to verify if an even better prediction is achieved.

For DL models, others techniques such as Hilbert Huang Transform (HHT) could be applied to generate others spectrum images. HHT is highly related with EMD, which, based on the ML results, could be a suitable choice. Other possibility is to work on more complex CNN architectures, adding deeper and different layers, without forgetting it would intensify the computational cost.

4.3 DISCUSSION

Some remarks could be made when comparing ML and DL methodologies to solve the same problem of prediction RUL from bearings based on a vibration dataset. As expected, the extraction of relevant features is necessary to be done and, in the ML approach, it was performed by the steps of feature selection, data sampling and the application of EMD algorithm. The use of those techniques improved the quality of data and provided a great prediction. However, it is necessary to highlight the difficulty in defining many parameters of each previous step (e.g. which representative measure for feature selection, the frequency rate in data sampling and the regressors used for EMD).

Conversely, DL approach transfers all complex preprocessing steps to the learning model and, to understand the degradation behavior, huge quantity of data is often positive. The use of CNN requires only adaptation of the raw input (i.e. time series vibration data) into an appropriate form (i.e. spectrum images). Also, for this application, the easier-to-create DL models presented results comparable to the best ML models, which is a hint for its exploration.

A current ongoing research analyze these methodologies for other bearings provided by the IEEE challenge. As initial results, the DL methodologies are proven to be way more suitable than the ML methodologies. Moreover, even with the pre-processing procedure pointed in section 4.2.1, handling data for many bearings was impracticable to generate an ML model. These preliminary results make sense when more data become available, which is particularly important for DL, and it is possible to use several bearings to create a general model easier than in the ML case. On the other hand, the simplicity provided by the DL application could not be seen as a gold solution. Actually, there is a necessity to defining a suitable architecture of the network, which surely impacts the performance, and to assure that the data, normally provided by automatic sensor, is actually prepared to be utilized.

5 CONSTRUCTION OF PERSONAL PROTECTIVE EQUIPMENT DETECTOR USING VIDEO IMAGES

Some results discussed on this chapter have been published on the European Safety and Reliability Conference (Trondheim) and presented on the *Encontro de Pesquisa e Pós-Graduação em Engenharia de Produção* (Florianópolis) in 2018.

5.1 CONTEXT

Even with the scientific and technological progress, statistics provided by the International Labor Organization (ILO) demonstrate that working conditions in many countries (e.g. European Union) have not changed to such a degree as to significantly reduce the problem of occupational injuries (Cavazza & Serpe, 2009). Therefore, every effort to decrease the number of accidents or, at least, maintain its rate at an acceptable range is highly important, and can be employed either by organizational actions, collective training or individual safeguard.

The traditional approach to avoid loss is the implementation of barriers, which plays a central role in the prevention of accidents. Sklet (2006) defines safety barriers as ‘physical and/or non-physical means planned to prevent, control, or mitigate undesired events or accidents.

Indeed, there are many opportunities to interrupt or change an accident sequence of events before it evolves into a loss. First, an answer is to change the preconditions for an accident to occur by eliminating the energy source or modifying the energy characteristics from the hazard. Second, barriers may interrupt, dilute, or redirect the energy flow during the latter part of the accident process (e.g. separating the victim from the energy flow). As a last barrier, it is possible to improve the victim’s ability to endure the energy flow (e.g. wearing some protective equipment), which is the ultimate protection to avoid damage (KJELLÉN; ALBRECHTSEN, 2017).

In this context, Personal Protective Equipment (PPE) is usually adopted to protect the individual against health or safety risks at work. It includes items related with protection of head, face, eye, hand, arms, and legs (Health and Safety Executive, 2013). There are consolidated regulations for the usage of PPE in industries (Occupational Safety and Health Administration - OSHA 2004; U.S. Homeland Security 2002) that aim to decrease the frequency of misuse or absence of PPE. Also, PPE’s positive impacts are very significant (e.g.

rate of eye injury and lost work time can be reduced by 50% or more when PPE is worn (Lipscomb, 2000)).

Head, as a vital body part containing possibly the most important human organ, needs appropriate attention. Every year, approximately 1.7 million people are hospitalized or die as a result of a traumatic brain injury (TBI) only in the United States (McCrory *et al.*, 2009). Protective headgear and helmets decrease the potential for severe TBI following a collision by reducing the acceleration of the head upon impact, thereby decreasing both the brain-skull collision, as well as the sudden deceleration induced axonal injury (Newman *et al.*, 2005).

There are several types for head protection such as industrial safety helmets, bump caps and firefighters' helmets. The use of those equipment is necessary in activities like low-level fixed objects with risk of collision (e.g. pipework, machines, scaffolding) and transport activities involving the risk of falling material (e.g. hoists, lifting plant, conveyors) (Health and Safety Executive, 2015).

The problem relies on the fact that, even with understanding about the safety improvement that the usage of PPE leads, its usage is often neglected in industry. The report of the ILO estimates that 2.34 million people die every year in the world due to occupational accidents, some of these deaths caused by non-use of PPE (INTERNATIONAL LABOUR OFFICE, 2011). A common approach is to impose fines and penalization to workers, who do not wear the required PPE when performing specific activities. However, supervision to guarantee its use is normally performed in person by a higher-level employee, which makes almost impossible to control all operators during the whole labor time.

Indeed, there is an extensive discussion concerning ethical issues in workplace surveillance, referring to management's ability to monitor, record and track employee performance, behaviors and personal characteristics in real time (BALL, 2010). Most of the discussion involves the so-called Electronic Performance Monitoring (EPM) about employee's control in social and technological forms (e.g. Internet and email monitoring, location tracking, biometrics) and the understanding of privacy boundaries surrounding employee information (Allen *et al.* 2015, Alder 1998). However, the discussion is to assure that proper safety protocol is followed, preventing injury to employees, as well as avoiding damage to the assets through a consistent and trustworthy model.

Thus, exploring similar tools and challenges to detect usage of PPE in order to avoid accidents in industries represents an interesting case and an automatic method for monitoring its usage is significantly worthy for industrial safety. Therefore, this chapter aims to develop a model for automatic PPE detection from industrial video streams using CV and DL for object detection, as well as an ML for object interaction.

5.2 REAL-TIME OBJECT DETECTION

As previously mentioned in section 3.2, CV studies the automated extraction of information from images and videos. The development of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms in the CV field (YILMAZ; JAVED; SHAH, 2006). In this context, visual object tracking has been constantly studied and presents three key steps for detection in video analysis: detection of movement of objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior (YILMAZ; JAVED; SHAH, 2006). Essentially, the basis of visual object tracking is to robustly estimate the motion state (i.e., location, orientation, size, etc.) of a target object in each frame of an input image sequence (LI *et al.*, 2013a).

Specifically, intelligent visual surveillance systems deal with the real-time monitoring of persistent and transient objects within a specific environment (Valera & Velastin, 2005). The goal of these systems is not only to put cameras in the place of human eyes, but create an entire surveillance system as automatically as possible (Hu *et al.*, 2004).

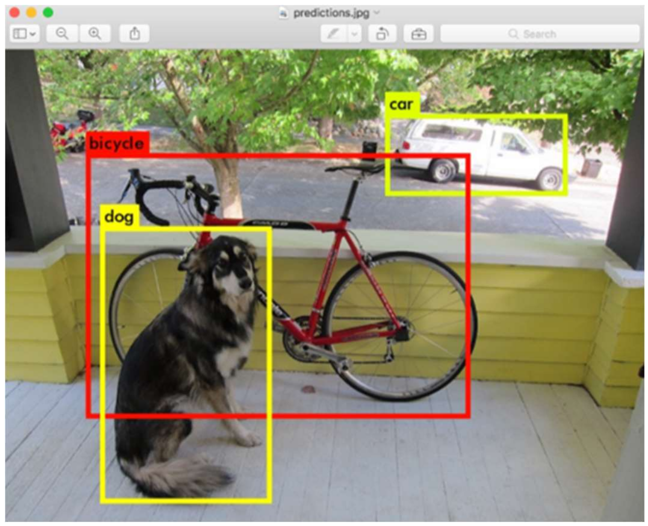
There exist some well-known visual surveillance systems such as W4 (Haritaoglu *et al.*, 2000); Haar-wavelet Adaboost (Enzweiler & Gavrila, 2009) and ViBe (Barnich & Van Droogenbroeck, 2011), mainly developed to detect different vehicle types, groups of people, pedestrians, people access control. Every system is developed seeking to compensate the capability limitation of human operators in monitoring enormous number of cameras at the same time.

Techniques from statistical pattern recognition have, since the revival of NNs, obtained a widespread use in digital image processing (Egmont-Petersen *et al.*, 2002). As already characterized in section 2.4, CNN has been successfully studied in fields such as speech recognition (Hinton *et al.*, 2012), vibration analysis (Guo *et al.*, 2016), electronic nose data (Längkvist *et al.*, 2013) and physiological data (Mirowski *et al.*, 2008). However, surely, the most promising results are found in the field of CV, bringing impressive developing in tasks like automatic object and face recognition.

One promising, open and free project that uses CNN for object detection is You only look once (YOLO). YOLO is a system for detecting objects and was first created on the Pascal

VOC 2012 dataset, detecting the 20 Pascal object classes, such as person, birds, dogs, car, bicycle, bottle, table and chair, as can be seen in Figure 17 (Redmon *et al.*, 2015)

Figure 17 - Example of object detection using YOLO.



Source: Adapted from Redmon *et al*. (2015).

The developers adopted a different approach than the standard object detection models that uses classifier based-systems applied at multiple locations and scales in an image, which typically considers as detections high scored regions of the image. In YOLO, a single CNN is executed to the full image in an architecture constituted of 24 convolutional layers followed by 2 fully connected layers. This network divides the image into regions and predicts bounding boxes and probabilities for each region, with these bounding boxes being weighted by the predicted probabilities. It has considerable advantages over other object detection models, once it looks at the whole image, and then its predictions are informed by global context in the image (Redmon *et al.*, 2015).

Still, an improved model, YOLOv2, has already been developed. More robust, detecting more than 9000 objects without losing real-time performance, YOLOv2 is a state-of-art object detection system with comparable or with even better results than many other systems (Redmon & Farhadi, 2017). Moreover, YOLO project allows inclusion of objects that were not on its detection basis, supporting training of a new model, allowing adaptation for different purposes.

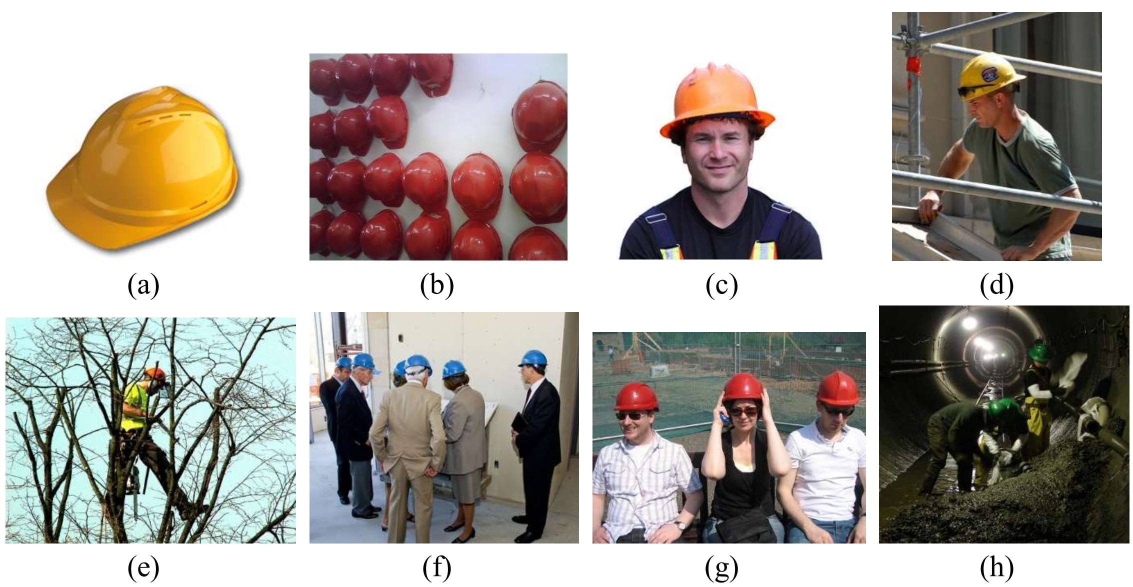
5.3 APPLICATION

In this section, YOLOv2 is used as a key tool to develop a new model to automatically detect PPE usage. Specifically, the goal is to identify whether workers were properly wearing a safety helmet when performing some activities in which the protection was required and/or mandatory.

5.3.1 Helmet detection using YOLO

YOLO project easily provides a pre-trained model, which could be used as a basis for detecting new types of objects. As any AI algorithm, YOLO requires a training dataset that will ‘teach’ the machine how an unknown object looks like. In this case, the desired PPE (i.e. hard helmets) is not a ‘known’ object for YOLO and, hence, it is necessary to create a database. For the specific goal, 731 images containing helmets were used to give sufficient information about its appearance. Distinct situations (Figure 18) were considered such as (a) isolated helmets, (b) multiple helmets, (c) isolated person with helmed, (d) backgrounded person with helmed, (e) covered person with helmet, (f) indoor multiple people wearing helmets, (g) outdoor multiple people wearing helmets and (h) low light exposure, for example.

Figure 18 - Example of distinct helmet images used for training

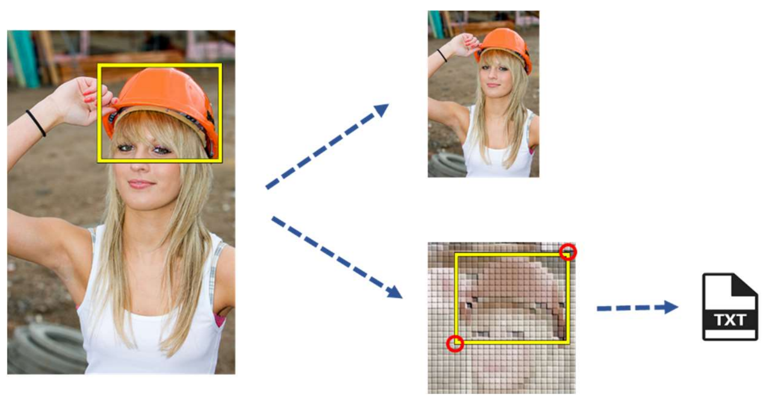


Source: The Author (2020)

All images were collected from ImageNet (JIA DENG *et al.*, 2009), an image database organized according to the WordNet (FELLBAUM, 1998) hierarchy, in which each node of the hierarchy is depicted by hundreds and thousands of images. It presents useful resource for researchers that need image data, containing many classes of items.

The location of helmets on images, which is the goal for the PPE detection, were annotated manually for each of the 731 pictures (i.e. bounding boxes for all imagens were constructed). The location problem is trickier than the classification problem: in the latter, the goal is to identify only whether an object appears in the image, while the former aims to define exactly where (i.e. which pixels) it appears. In other words, the location problem could be seen as a further step after the classification problem. Hence, to the training process, not only the image itself is stored, but also the coordinate of the bounding boxes (Figure 19).

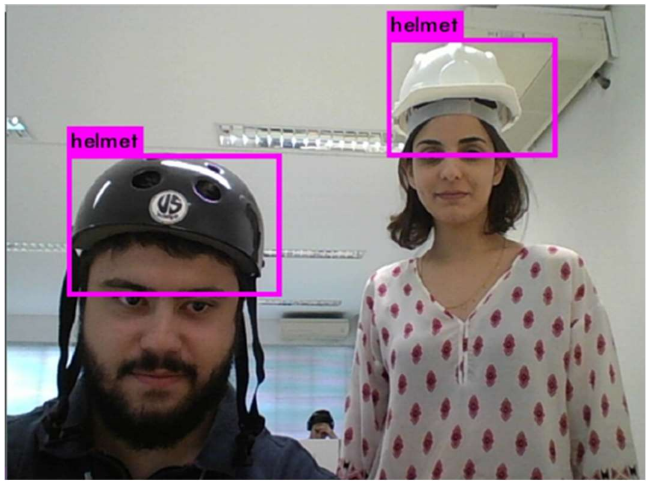
Figure 19 - Database generated for helmet detection in YOLO. Besides the image (above), bounding boxing coordination (red circles below) are also stored.



Source: The Author (2020)

After feed YOLO pre-trained network with the PPE images and its respective bounding boxes files, the detection model could be created. The network was trained for about 8 hours, running in a Nvidia GeForce GTX 960m GPU, with 4GB of video random access memory (VRAM). Once the algorithm finished its training, the helmet detection model could be applied to a specific image or to a video stream, such as a camera feed, processing every frame. The model runs in real-time, maintaining the frame rate of the camera (30 frames per second – FPS). Figure 20 depicts the model applied to a standard web camera video streaming.

Figure 20 - Helmet detection model applied for video streaming.



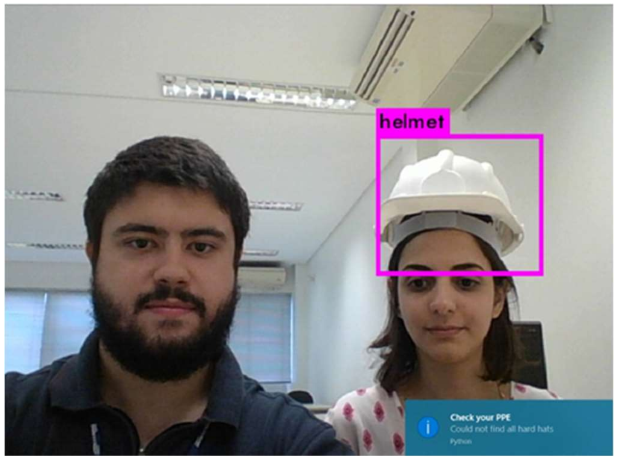
Source: The Author (2020)

5.3.2 Validation, warnings and support decisions

Then, a script was created to alert surveillance operators whether an abnormal situation appears (i.e. helmets were not detected). As a first validation procedure, a simple experiment consisted in a room containing a specific number of employees that should be using helmets was performed. A video camera (i.e. simulating a surveillance camera), recorded the room during few minutes and, for each frame, the algorithm detects the use of helmet and counts how many are present in the scene. If the number of detected helmets is different from the number of previously defined people in the room for more than a brief period (e.g. 10 seconds, or 30 seconds), then an alert was emitted.

Moreover, if the number of detected helmets does not return to its desired value, alerts may continue to be emitted in a previously defined interval (e.g. 30 seconds, or two minutes). Properly adjusting this period avoids unnecessary alerts after recognition the anomaly situation. Figure 21 shows a computer screen when an alert was presented (i.e. one person is not using helmet in the image).

Figure 21 - Alert emitted when model detect anomaly situation.



Source: The Author (2020)

However, in this first validation procedure, even though location values (i.e. bounding boxing) are generated in a relatively correct position, this information was not used for warning emission. It means, for example, that if an operator was wrongly using the helmet (i.e. holding helmet in hand; not wearing) but it appears in the image, the script would not warn.

Hence, a second methodology was proposed. In this case, multiple detections were carried out (i.e. detection of helmet and person), such as in Figure 22. In this case, the goal is to verify if the helmet is actually placed over the head of a person. The methodology consists in locating the bounding boxes of the person and the helmet, and calculate the distance between them. Specifically, the center of the upper bound of the person should be close to the upper bound of the helmet. Obviously, it is necessary to have at least the same number of helmets, when compared with the number of people on the picture. The computation of a simple threshold to evaluate the maximum distance between both bounding boxes (i.e. person and helmet) to detect the usage of PPE did not prove to be reliable, and, hence, a refined approach was used.

Figure 22 - Multiple detection (helmet and person) provided by the model

Source: The Author (2020)

To that end, a classification SVM is applied to evaluate which relation (i.e. distance) should be considered as a use/not use of helmet. 135 images (and the bounding boxing distance) were manually selected to train the algorithm while 54 images where used for test purposes. Once again, RBF was used as the kernel function, here with $C = 1$ and $\gamma = 0.2$. The performance of this methodology to correctly identify the usage of PPE is provided in Table 9, in which '0' is 'no usage','1' denotes 'usage' and 'Acc. (%)' represents the accuracy in percentage.

Table 9 - Accuracy of correct use of PPE

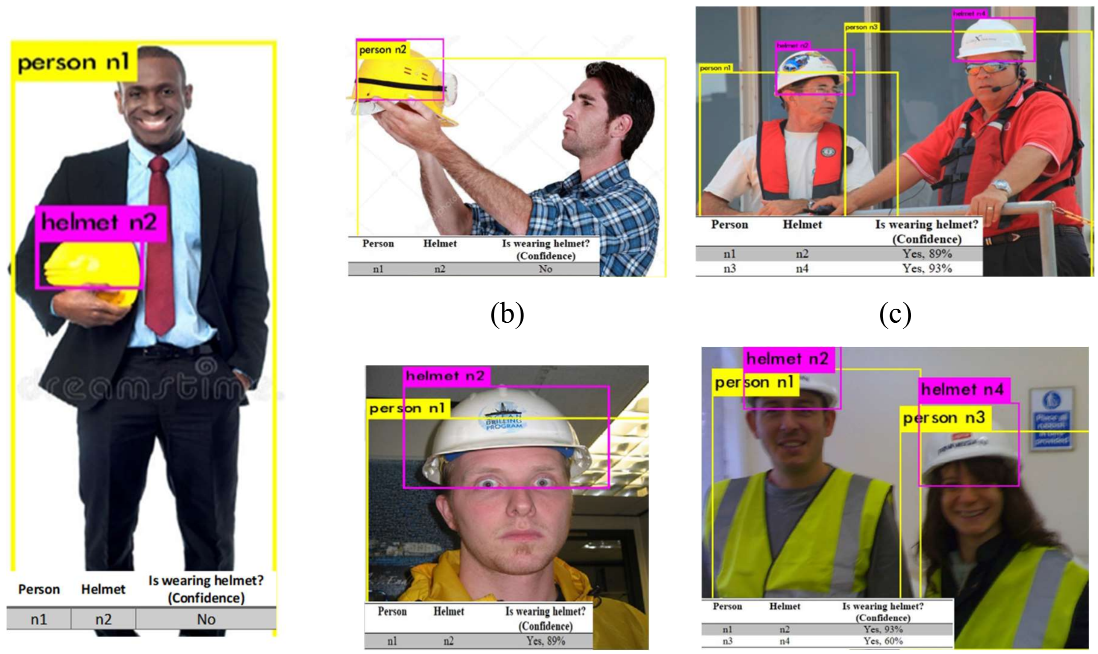
		True Label					
		Training			*Test*		
		0	1	Acc. (%)	0	1	Avg (%).
Predicted Label	0	80	8	90.37	34	1	90.44
	1	5	42		2	17	

Source: The Author (2020)

Note that the above performance is not related with the PPE detection, but rather with the correct usage of it. Hence, the model presented relatively robust ability, especially in detecting a 'no use' of PPE, which is the worst case on safety environments. Figure 23 also illustrates the proposed algorithm working and correctly predicting the usage (or not) of PPE.

In (a) and (b), the algorithm correctly warns a no usage of helmet, while in (c), (d) and (e), the desired usage is verified for one or more people.

Figure 23 - PPE usage detection based on SVM.



(a) (b) (c) (d) (e)

Source: The Author (2020)

In practice, the developed model could be explored as a tool in different contexts, supporting decisions for the safety manager. Other types of warnings (e.g. depending on how long operators remain without PPE; how many operators are not wearing the PPE) could be implemented and customable, providing information for the decision-maker to determine whether or not someone must be notified. Moreover, it is also possible to use information and statistics provided by the model (e.g. how many times alerts were displayed per day; how long operators had remained without PPE) as a safety indicator.

5.4 DISCUSSION

This chapter presented an approach for automatically detecting PPE usage in distinct environments, using object detection with YOLO. By using YOLO, this method achieves a reasonable balance between speed and confidence, running the PPE detection in real-time, which results in relatively low computational resource usage. Moreover, SVM were used to relate two detections (i.e. helmet and person) in order to identify if the PPE is correctly placed. The developed model could lead to beneficial results to safety engineering since the detection is performed automatically and does not require constant human attention. With wider adaptations, this surveillance technology could be also implemented to monitor other barriers than PPE. For example, a similar image-based methodology could act as a redundancy to other sensors to improve detection of hazards (e.g. fire, toxic gases) in order to interrupt the energy flow in case of accidents.

However, some comments on this PPE detection model should be done, which is an ongoing work. Despite working well on images, the SVM model (i.e. to determine the right position of PPE) is not implemented to process videos in real-time. Moreover, for a complete validation, the usage of a public dataset of PPE wearing (if available) would be extremely valuable. A further step is to use real surveillance videos as input, detecting the usage of PPE in realistic environment, preventing accidents and providing an improvement on the safety monitoring system of industries.

As a future research, this model could be extended for application such as identification of other types of PPEs and simultaneously monitoring its usage. Still, implementation of a real time alert for the operators (e.g. a particular warning light is lit somewhere in the room) connected with the model would emphasize (or create) the sense of autoregulation among them, reducing (or sharing) the surveilling workload expected for the supervisor.

6 FINAL CONSIDERATIONS

Over the years, several AI approaches have been proposed in a diversity of applications of real-world problems. Specifically, ML and DL algorithms have provided successful models in engineering systems and are the state-of-art methodologies for almost every environment. This work presented three specific applications in which AI models were developed to solve problems in risk and reliability context.

In the first application, a real-time drowsiness detection model was proposed based on video streams of web cameras. The model used an ML approach not requiring GPU, which could be used in almost every standard computer. The model relied in neuroscience rules to warn about drowsiness and was validated in a public database with videos of subjects performing drowsiness experiments. In this case, the use of ML is justified considering low-cost computers or the easier linkage that ML implementations have if one intends to attach the drowsiness model to other systems (e.g. other types of cameras).

In the second application, ML and DL approaches were considered to estimate RUL from bearings based on its vibration signal. As expected, ML approach required a lot of human effort in analysing pre-processing steps when compared with DL approach. However, if those stages are performed adequately, the model provides great results. In contrast, if one has available computational force and numerous data, DL approach seems more suitable to derive general models.

In the third application, a real-time PPE detection was created based on YOLO, an open and free DL project for object detection. Here, a new database had to be created to teach YOLO's pre trained model how to recognize the shape of the specific object (i.e. hard hat). Moreover, alert systems were created to automatically warn if pre-defined conditions are met and, based on SVM, if PPE is properly worn.

In all the three contexts, AI methods are used to support decision makers with relative success, however, on the system level, many other variables should be noticed. Hence, the presented approaches and the obtained models have to be preliminary investigated to be used in similar situations. The data-driven approaches rely only in the data and ignore all the engineering theory in the creation of models may bias fundamental aspects. Moreover, the availability and accessibility dataset narrow the use of these methodologies, and these methods cannot just be used in some situations. Hence, each problem has to be treated as a particular case to analyze the best approach to be used.

The revolution introduced by ML and extended by DL is still happening and more sophisticated methods and algorithms will surely appear. Although those models are used as support today, the level of autonomy will grow exponentially in the next few years and how to maximize the benefits brought to society by these tireless computational workers.

REFERENCES

ÅKERSTEDT, Torbjörn; ANUND, Anna; AXELSSON, John; KECKLUND, Göran. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of Sleep Research*, v. 23, n. 3, p. 240–252, 2014.

ÅKERSTEDT, Torbjörn; GILLBERG, Mats. Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, v. 52, n. 1–2, p. 29–37, 1990.

AL-FUQAHA, Ala; GUIZANI, Mohsen; MOHAMMADI, Mehdi; ALEDHARI, Mohammed; AYYASH, Moussa. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys and Tutorials*, v.17, n. 4, p. 2347 - 2376, 2015.

ALDER, G. Stoney. Ethical Issues in Electronic Monitoring: A Consideration of Deontological Perspectives. *Journal of Business Ethics*, v. 17, n. 7, p. 729–743, 1998.

ALLEN, Myria Watkins; COOPMAN, Stephanie J.; HART, Joy L.; WALKER, Kasey L. Workplace Surveillance and Managing Privacy Boundaries. *Labor History*, v. 51, n. 1, p. 172–200, 2015.

AMBHORE, Nitin; KAMBLE, Dinesh; CHINCHANIKAR, Satish; WAYAL, Vishal. Tool condition monitoring system: A review. *Materials Today: Proceedings*, v. 2, n. 4–5, p. 3419–3428, 2015.

ANGUITA, Davide; GHIO, Alessandro; ONETO, Luca; PARRA, Xavier; REYES-ORTIZ, Jorge L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. *In*: 2012, *Lecture Notes in Computer Science.* p. 216–223.

ANOKHIN, Andrey P.; HEATH, Andrew C.; MYERS, Erin; RALANO, Angela; WOOD, Suzanne. Genetic influences on prepulse inhibition of startle reflex in humans. *Neuroscience Letters*, 2003.

ANSIAU, D.; WILD, P.; NIEZBORALA, M.; ROUCH, I.; MARQUIÉ, J. C. Effects of working conditions and sleep of the previous day on cognitive performance. *Applied Ergonomics*, v. 39, n. 1, p. 99–106, 2008.

BABENKO, Boris; MING-HSUAN YANG; BELONGIE, Serge. Robust Object Tracking with Online Multiple Instance Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 33, n. 8, p. 1619–1632, 2011.

BACIVAROV, Ioana; IONITA, Mircea; CORCORAN, Peter. Statistical models of appearance for eye tracking and eye-blink detection and measurement. *IEEE Transactions on Consumer Electronics*, 2008.

BALL, Kirstie. Workplace surveillance: An overview. *Labor History*, v. 51, n. 1, p. 87–106, 2010.

BARNICH, Olivier; VAN DROOGENBROECK, Marc. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, v. 20, n. 6, p. 1709–1724, 2011.

BATISTA, Jorge P. Locating Facial Features using an Anthropometric Face Model for Determining the Gaze of Faces in Image Sequences. *Image Analysis and Recognition, ICIAR*, 2007.

BECERRA-RIERA, Fabiola; MORALES-GONZÁLEZ, Annette; MÉNDEZ-VÁZQUEZ, Heydi. Facial marks for improving face recognition. *Pattern Recognition Letters*, v. 113, p. 3–9, 2017.

BELENKY, Gregory et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, 2003.

BEN-DAVID, Shai; SHALEV-SHWARTZ, Shai. *Understanding Machine Learning: From Theory to Algorithms*. : Cambridge University Press, 2014. .

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1798–1828, 2013.

BHAVSAR, Punitkumar; SRINIVASAN, Babji; SRINIVASAN, Rajagopalan. Quantifying situation awareness of control room operators using eye-gaze behavior. *Computers and Chemical Engineering*, 2017.

BORGHINI, Gianluca; ASTOLFI, Laura; VECCHIATO, Giovanni; MATTIA, Donatella; BABILONI, Fabio. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, v. 44, p. 58–75, 2014.

BOŠKOSKI, Pavle; GAŠPERIN, Matej; PETELIN, Dejan; JURIČIĆ, Dani. Bearing fault prognostics using Rényi entropy based features and Gaussian process models. *Mechanical Systems and Signal Processing*, 2015.

BOUSDEKIS, Alexandros; MAGOUTAS, Babis; APOSTOLOU, Dimitris; MENTZAS, Gregoris. Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *Journal of Intelligent Manufacturing*, 2015.

BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, 2012.

BRUNI, Renato; BIANCHI, Gianpiero. Website categorization: A formal approach and robustness analysis in the case of e-commerce detection. *Expert Systems with Applications*, 2020.

CAFFIER, Philipp P.; ERDMANN, Udo; ULLSPERGER, Peter. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, v.

89, n. 3, p. 319–325, 2003.

CARCAGNÌ, Pierluigi; DEL COCO, Marco; LEO, Marco; DISTANTE, Cosimo. Facial expression recognition and histograms of oriented gradients: a comprehensive study. *SpringerPlus*, 2015.

CAVAZZA, Nicoletta; SERPE, Alessandra. Effects of safety climate on safety norm violations: exploring the mediating role of attitudinal ambivalence toward personal protective equipment. *Journal of Safety Research*, v. 40, n. 4, p. 277–283, 2009.

CAVUSCULU, Melih; YETIK, Imam Samil; YEGINER, Mete. Fully automated blink detection for uveal melanoma radiotherapy. *In*: 2017, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. p. 628–631.

CHANG, Long-yi; CHUNG, Yi-nung; LIN, Chia-hung; CHEN, Jian-liung; KUO, Chao-lin; CHEN, Shi-jaw. Mechanical Vibration Fault Detection for Turbine Generator Using Frequency Spectral Data and Machine Learning Model : Feasibility Study of Big Data Analysis. *Sensors and Materials*, v. 30, n. 4, p. 821–832, 2018.

CHEN, Jinglong et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, v. 70–71, p. Pages 1-35, 2016.

CHEN, Jonathan H.; ASCH, Steven M. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. *New England Journal of Medicine*, 2017.

CHEN, Xianming; WANG, Tieliu; DING, Mingming; WANG, Jing; CHEN, Jianqing; YAN, Jun Xia. Analysis and prediction on the cutting process of constrained damping boring bars based on PSO-BP neural network model. *Journal of Vibroengineering*, v. 19, n. 2, p. 878–893, 2017.

CHEN, Xue Wen; LIN, Xiaotong; XUE-WEN CHEN; XIAOTONG LIN. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, v. 2, p. 514–525, 2014.

CHENG, Bing; TITTERINGTON, D. M. Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 1994.

CIREŞAN, Dan; MEIER, Ueli; MASCI, Jonathan; SCHMIDHUBER, Jürgen. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 2012.

CLAESEN, Marc; SIMM, Jaak; JUMUTC, Vilen. *Optunity*. Leuven: *[s. n.]*, 2015. Disponível em: http://www.optunity.net

COFRE-MARTEL, Sergio; KOBRICH, Philip; DROGUETT, Enrique Lopez; MERUANE, Viviana. Deep Convolutional Neural Network-Based Structural Damage Localization and Quantification Using Transmissibility Data. *Shock and Vibration*, 2019.
COGLIANESE, Cary; LEHR, David. Regulating by robot: administrative decision making in the machine-learning. *Georgetown Law Journal*, v. 105, p. 1147, 2016.

COHEN, Michael X. A better way to define and describe Morlet wavelets for time-frequency analysis. *NeuroImage*, 2019.

COLE, D. F. *Adler's Physiology of the Eye: Clinical Application*.

DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. *In*: 2005, *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005.* p. 886–893.

DEMENTYEV, Artem; HOLZ, Christian. DualBlink: AWearable Device to Continuously Detect, Track, and Actuate Blinking For Alleviating Dry Eyes and Computer Vision Syndrome. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017.

DING, Michael E. et al. Measurement of resident fatigue using rapid number naming. *Journal of the Neurological Sciences*, 2019.

DINGES, David F.; POWELL, John W. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, v. 17, n. 6, p. 652–655, 1985.

DIVJAK, M.; BISCHOF, H. Eye Blink Based Fatigue Detection for Prevention of Computer Vision Syndrome. *Conference on Machine Vision Applications*, p. 350–353, 2009. Disponível em: http://www.mva-org.jp/Proceedings/2009CD/papers/10-04.pdf

DROGUETT, E. L.; LINS, I. D.; MOURA, M. C.; ZIO, E.; JACINTO, C. M. Variable selection and uncertainty analysis of scale growth rate under pre-salt oil wells conditions using support vector regression. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, v. 229, n. 4, p. 319–326, 2014.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks—a review. *Pattern Recognition*, v. 35, n. 10, p. 2279–2301, 2002.

EL-THALJI, Idriss; JANTUNEN, Erkki. A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, v. 60, p. 252–272, 2015.

ENZWEILER, Markus; GAVRILA, Dariu M. Monocular pedestrian detection: Survey and experiments. *In*: 2009, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. p. 2179–2195.

ESPINOSA, Julián; ROIG, Ana Belén; PÉREZ, Jorge; MAS, David. A high-resolution binocular video-oculography system: Assessment of pupillary light reflex and detection of an early incomplete blink and an upward eye movement. *BioMedical Engineering Online*, 2015.
FELLBAUM, Christiane. *WordNet: An Electronic Lexical Database*. Cambridge: MIT press, 1998. v. 71 .

FERNANDEZ, Alberto; USAMENTIAGA, Ruben; CARUS, Juan Luis; CASADO, Ruben.

Driver distraction using visual-based sensors and algorithms. *Sensors (Switzerland)*, v. 16, n. 11, 2016.

FLORES, Marco Javier; ARMINGOL, José María; DE LA ESCALERA, Arturo. Real-time warning system for driver drowsiness detection using visual information. *Journal of Intelligent and Robotic Systems: Theory and Applications*, v. 59, n. 2, p. 103–125, 2010.

FRANÇOIS, Clémentine; HOYOUX, Thomas; LANGOHR, Thomas; WERTZ, Jérôme; VERLY, Jacques G. Tests of a new drowsiness characterization and monitoring system based on ocular parameters. *International Journal of Environmental Research and Public Health*, 2016.

FRED AGARAP, Abien M. *Deep Learning using Rectified Linear Units (ReLU)*. . Disponível em: https://github.com/AFAgarap/relu-classifier. Acesso em: 14 maio. 2019.

FUMEO, Emanuele; ONETO, Luca; ANGUITA, Davide. Condition based maintenance in railway transportation systems based on big data streaming analysis. *In*: 2015, *Procedia Computer Science.*

GARCÍA NIETO, P. J.; GARCÍA-GONZALO, E.; SÁNCHEZ LASHERAS, F.; DE COS JUEZ, F. J. Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. *Reliability Engineering and System Safety*, v. 138, p. 219–231, 2015.

GONZÁLEZ GARCÍA, Cristian; MEANA-LLORIÁN, Daniel; PELAYO G-BUSTELO, B. Cristina; CUEVA LOVELLE, Juan Manuel; GARCIA-FERNANDEZ, Nestor. Midgar: Detection of people through computer vision in the Internet of Things scenarios to improve the security in Smart Cities, Smart Towns, and Smart Homes. *Future Generation Computer Systems*, 2017.

GOODFELLOW, I. J. et al. Generative Adversarial Nets. *In*: 2016, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.*

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. : MIT press, 2016. .

GOYAL, D.; PABLA, B. S. The Vibration Monitoring Methods and Signal Processing Techniques for Structural Health Monitoring: A Review. *Archives of Computational Methods in Engineering*, 2016.

GU, Wang Huan; ZHU, Yu; CHEN, Xu Dong; HE, Lin Fei; ZHENG, Bing Bing. Hierarchical CNN-based real-time fatigue detection system by visual-based technologies using MSP model. *IET Image Processing*, v. 12, n. 12, p. 2319–2329, 2018.

GUO, Liang; LI, Naipeng; JIA, Feng; LEI, Yaguo; LIN, Jing. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 2017.

GUO, Xiaojie; CHEN, Liang; SHEN, Changqing. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement: Journal of the*

International Measurement Confederation, v. 93, p. 490–502, 2016.

HADDADIN, S. Towards the robotic co-worker. *Springer Tracts in Advanced Robotics*, v. 90, p. 195–215, 2014.

HARITAOGLU, Ismail; HARWOOD, David; DAVIS, Larry S. W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, p. 809–830, 2000.

HEALTH AND SAFETY EXECUTIVE. *Personal protective equipment (PPE) at work: A brief guide*. . . Disponível em: http://www.hse.gov.uk/pubns/indg174.pdf

HEALTH AND SAFETY EXECUTIVE. *Personal protective equipment at work*. . v. 25 .

HINTON, Geoffrey et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, v. 29, n. 6, p. 82–97, 2012.

HOLLANDER, Myles; WOLFE, Douglas A.; CHICKEN, Eric. *Nonparametric Statistical Methods*. Third ed. New Jersey: John Wiley & Sons, 2014. .

HU, Weiming; TAN, Tieniu; WANG, Liang; MAYBANK, Steve. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, v. 34, n. 3, p. 334–352, 2004.

HUANG, Bin; DI, Yuan; JIN, Chao; LEE, Jay. Review of Data-Driven Prognostics and Health Management Techniques: Lessions Learned From Phm Data Challenge Competitions. *Machine Failure Prevention Technology 2017*, 2017.

HUANG, Shengzhi; CHANG, Jianxia; HUANG, Qiang; CHEN, Yutong. Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, v. 511, p. 764–775, 2014.

INGRE, Michael; ÅKERSTEDT, Torbjörn; PETERS, Björn; ANUND, Anna; KECKLUND, Göran. Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of Sleep Research*, 2006.

INTERNATIONAL LABOUR OFFICE. *ILO introductory report: global trends and challenges on occupational safety and health: XIX World Congress on Safety and Health at Work*. .

JACOBÉ DE NAUROIS, Charlotte; BOURDIN, Christophe; BOUGARD, Clément; VERCHER, Jean Louis. Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness. *Accident Analysis and Prevention*, 2018.

JAIN, Anil K.; MAO, Jianchang; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. *Computer*, v. 29, n. 3, 1996.

JAIN, Siddhant; JAIN, Abhishek; VERMA, Shikhar; SUSAN, Seba; SHARMA, Aakash.

Fuzzy match index for scale-invariant feature transform (SIFT) features with application to face recognition with weak supervision. *IET Image Processing*, 2015.

JÄKEL, Frank; SCHÖLKOPF, Bernhard; WICHMANN, Felix A. A tutorial on kernel methods for categorization. *Journal of Mathematical Psychology*, v. 51, n. 6, p. 343–358, 2007.

JAN ERIK SOLEM. Programming Computer Vision with Python. *Programming Computer Vision with Python*, p. 264, 2012.

JENSEN, Oh; LARSEN, R. *Implementing the viola-jones face detection algorithm*. . .

JIA DENG; WEI DONG; SOCHER, R.; LI-JIA LI; KAI LI; LI FEI-FEI. ImageNet: A large-scale hierarchical image database. *In*: 2009, *2009 IEEE Conference on Computer Vision and Pattern Recognition.* p. 248–255.

JIANG, Peng; MAGHREBI, Mojtaba; CROSKY, Alan; SAYDAM, Serkan. Unsupervised Deep Learning for Data-Driven Reliability and Risk Analysis of Engineered Systems. *In*: *Handbook of Neural Computation*. : Academic Press, 2017. p. 417–431. .

JIMENEZ-PINTO, J.; TORRES-TORRITI, M. Face salient points and eyes tracking for robust drowsiness detection. *Robotica*, 2012.

JOHN, Sofia Jennifer; SHARMILA, Sree T. Real time blink recognition from various head pose using single eye. *Multimedia Tools and Applications*, 2018.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, v. 349, n. 6245, p. 255–260, 2015.

JUEFEI-XU, Felix; SAVVIDES, Marios. Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on NIST's face recognition grand challenge. *IEEE Transactions on Image Processing*, 2014.

KAIDA, Kosuke et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 2006.

KANG, Hang Bong. Various approaches for driver and driving behavior monitoring: A review. *In*: 2013, *Proceedings of the IEEE International Conference on Computer Vision.*

KAPLAN, Sinan; GUVENSAN, Mehmet Amac; YAVUZ, Ali Gokhan; KARALURT, Yasin. Driver Behavior Analysis for Safe Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 2015.

KAZEMI, Vahid; SULLIVAN, Josephine. One millisecond face alignment with an ensemble of regression trees. *In*: 2014, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. p. 1867–1874.

KHAN, Samir; YAIRI, Takehisa. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, v. 107, p. 241–265, 2018.

KIM, Ki Wan; HONG, Hyung Gil; NAM, Gi Pyo; PARK, Kang Ryoung. A study of deep CNN-based classification of open and closed eyes using a visible light camera sensor. *Sensors (Switzerland)*, 2017.

KING, Davis. *Dlib*. , 2003. Disponível em: http://dlib.net. Acesso em: 20 jul. 2019.

KJELLÉN, Urban; ALBRECHTSEN, Eirik. *Prevention of Accidents and Unwanted Occurrences*. : CRC Press, 2017. .

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 2006.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, p. 1–9, 2012.

LÄNGKVIST, Martin; CORADESCHI, Silvia; LOUTFI, Amy; BALAGURU RAYAPPAN, John Bosco. Fast classification of meat spoilage markers using nanostructured ZnO thin films and unsupervised feature learning. *Sensors (Switzerland)*, v. 13, n. 2, p. 1578–1592, 2013.

LANGKVIST, Martin; KARLSSON, Lars; LOUTFI, Amy. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, v. 42, n. 1, p. 11–24, 2014.

LASI, Heiner; FETTKE, Peter; KEMPER, Hans Georg; FELD, Thomas; HOFFMANN, Michael. Industry 4.0. *Business and Information Systems Engineering*, 2014.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015.

LEE, Jay; KAO, Hung An; YANG, Shanhu. Service innovation and smart analytics for Industry 4.0 and big data environment. *In*: 2014, *Procedia CIRP*.

LEE, Jong Jae; YUN, Chung Bang. Damage diagnosis of steel girder bridges using ambient vibration data. *Engineering Structures*, 2006.

LEE, Won Oh; LEE, Eui Chul; PARK, Kang Ryoung. Blink detection robust to various facial poses. *Journal of Neuroscience Methods*, v. 193, n. 2, p. 356–372, 2010.

LEFKOVITS, Szidonia; LEFKOVITS, Laszlo; EMERICH, Simina. Detecting the Eye and its Openness with Gabor Filters. *2017 5th International Symposium on Digital Forensic and Security (ISDFS)*, 2017.

LI, Xi; HU, Weiming; SHEN, Chunhua; ZHANG, Zhongfei; DICK, Anthony; HENGEL, Anton Van Den. A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology*, v. 4, n. 4, p. 1–48, 2013 a.

LI, Yang; JIANG, Xiaoyan; HWANG, Jenq Neng. Effective person re-identification by self-attention model guided feature learning. *Knowledge-Based Systems*, v. 187, p. 104832, 2020.

LI, Yongqiang; WANG, Shangfei; ZHAO, Yongping; JI, Qiang. Simultaneous facial feature tracking and facial expression recognition. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, v. 22, n. 7, p. 2559–2573, 2013 b.

LIAO, Linxia; KÖTTIG, Felix. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, 2014.

LIM, Julian; DINGES, David F. Sleep deprivation and vigilant attention. *In*: 2008, *Annals of the New York Academy of Sciences.*

LIN, Hsiung Cheng; YE, Yu Chen. Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms. *Advances in Mechanical Engineering*, 2019.

LINS, Isis Didier; ARAUJO, Moacyr; DAS, Márcio; MOURA, Chagas; SILVA, Marcus André; DROGUETT, Enrique López. Computational Statistics and Data Analysis Prediction of sea surface temperature in the tropical Atlantic by support vector machines. *Computational Statistics and Data Analysis*, v. 61, p. 187–198, 2013.

LINS, Isis; MOURA, M.; DROGUETT, E. Failure prediction of oil wells by support vector regression with variable selection, hyperparameter tuning and uncertainty analysis. *Chemical Engineering Transactions*, v. 33, p. 817–822, 2013.

LIPSCOMB, H. J. Effectiveness of interventions to prevent work-related eye injuries. *American journal of preventive medicine*, v. 18, n. 4 Suppl, p. 27–32, 2000. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/10793278

LIU, Jianzheng; LI, Jie; LI, Weifeng; WU, Jiansheng. Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 115, p. 134–142, 2016.

LIU, Mingfei; WU, Wei; GU, Zhenghui; YU, Zhuliang; QI, Fei Fei; LI, Yuanqing. Deep learning based on Batch Normalization for P300 signal detection. *Neurocomputing*, 2018.

LU, Yang. Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, v. 6, p. 1–10, 2017.

MAIOR, C. B. S.; MOURA, M. C.; LINS, I. D.; DROGUETT, E. L.; DINIZ, H. H. L. Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine. *IEEE Latin America Transactions*, v. 14, n. 11, p. 4603–4610, 2016.

MAIOR, C. B. S. et al. Real-time SVM classification for drowsiness detection using eye aspect ratio. *PSAM 2018 - Probabilistic Safety Assessment and Management*, n. September, 2018.

MAIOR, Caio Bezerra Souto; MOURA, Marcio das Chagas; LINS, Isis Didier. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. *Eksploatacja i Niezawodnosc - Maintenance and Reliability*, v. 21, n. 4, p. 610–619, 2019.

MAKEIG, Scott; INLOW, Mark. Lapse in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalography and Clinical Neurophysiology*, 1993.

MALLAT, Stephane. *A Wavelet Tour of Signal Processing*. . .

MAO, Wentao; HE, Jianliang; TANG, Jiamei; LI, Yuan. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Advances in Mechanical Engineering*, 2018.

MARTYNENKO, Alex. Computer Vision for Real-Time Control in Drying. *Food Engineering Reviews*, v. 9, n. 2, p. 91–111, 2017.

MASSOZ, Quentin; LANGOHR, Thomas; FRANCOIS, Clémentine; VERLY, Jacques G. The ULg multimodality drowsiness database (called DROZY) and examples of use. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016.

MCCRORY, Paul et al. Consensus statement on concussion in sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *In*: 2009, **J***ournal of Athletic Training*. p. 434–448.

MCKINLEY, R. Andy; MCINTIRE, Lindsey K.; SCHMIDT, Regina; REPPERGER, Daniel W.; CALDWELL, John A. Evaluation of eye metrics as a detector of fatigue. *Human Factors*, 2011.

MIROWSKI, Piotr W.; LECUN, Yann; MADHAVAN, Deepak; KUZNIECKY, Ruben. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. *In*: 2008, *Proceedings of the 2008 IEEE Workshop on Machine Learning for Signal Processing, MLSP 2008*. p. 244–249.

MITTAL, Ajay; KUMAR, Kanika; DHAMIJA, Sarina; KAUR, Manvjeet. Head movement-based driver drowsiness detection: A review of state-of-art techniques. *In*: 2016, *Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016.*

MNIH, Volodymyr et al. Human-level control through deep reinforcement learning. *Nature*, 2015.

MNIH, Volodymyr et al. Playing Atari with Deep Reinforcement Learning. *IJCAI International Joint Conference on Artificial Intelligence*, 2016.

MOESLUND, Thomas B.; HILTON, Adrian; KRÜGER, Volker. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, v. 104, n. 2- 3 SPEC. ISS., p. 90–126, 2006.

MORLET, J.; ARENS, G.; FOURGEAU, E.; GIARD, D. Wave propagation and sampling theory—Part II: Sampling theory and complex waves. *Geophysics*, v. 47, n. 2, p. 222–236, 1982.

MOURA, Márcio Das Chagas; ZIO, Enrico; LINS, Isis Didier; DROGUETT, Enrique. Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, v. 96, n. 11, p. 1527–1534, 2011.

NAGRANI, Arsha; CHUNG, Joon Son; XIE, Weidi; ZISSERMAN, Andrew. Voxceleb: Large-scale speaker verification in the wild. *Computer Speech and Language*, v. 60, 2020.

NECTOUX, Patrick et al. PRONOSTIA : An experimental platform for bearings accelerated degradation tests. *In*: 2012, *IEEE International Conference on Prognostics and Health Management, PHM'12.* p. 1–8. Disponível em: http://hal-obspm.ccsd.cnrs.fr/UNIV-BM/hal-00719503

NEWMAN, J. A.; BEUSENBERG, M. C.; SHEWCHENKO, N.; WITHNALL, C.; FOURNIER, E. Verification of biomechanical methods employed in a comprehensive study of mild traumatic brain injury and the effectiveness of American football helmets. *Journal of Biomechanics*, v. 38, n. 7, p. 1469–1481, 2005.

NICKEL, Maximilian; MURPHY, Kevin; TRESP, Volker; GABRILOVICH, Evgeniy. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, v. 104, n. 1, p. 11–33, 2016.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION (OSHA). **P***ersonal Protective Equipment* p. 1–32.

OMIDYEGANEH, Mona et al. Yawning Detection Using Embedded Smart Cameras. *IEEE Transactions on Instrumentation and Measurement*, v. 65, n. 3, p. 570–582, 2016.

PATIL, Meru A. et al. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. *Applied Energy*, v. 159, p. 285–297, 2015.

PEDREGOSA, Fabian; VAROQUAUX, G. *Scikit-learn: Machine learning in Python*. . v. 12 .

PEIRIS, M. T. R. et al. Identification of Vigilance Lapses using EEG/EOG by Expert Human Raters. *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, p. 5735–5737, 2005.

RAKSHITA, R. Communication Through Real-Time Video Oculography Using Face Landmark Detection. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, n. Icicct, p. 1094–1098, 2018.

RAMOS, Marilia Abílio; DROGUETT, Enrique López; MOSLEH, Ali; MOURA, Márcio das Chagas; MARTINS, Marcelo Ramos. Revisiting past refinery accidents from a human reliability analysis perspective: The BP Texas City and the Chevron Richmond accidents. *Canadian Journal of Chemical Engineering*, v. 95, n. 12, p. 2293–2305, 2017.

RANI, P. Sheela; SUBHASHREE, P.; DEVI, N. Sankari. Computer vision based gaze tracking for accident prevention. *In*: 2016, *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave).* : IEEE, 2016. p. 1–6.

RANZATO, Marc Aurelio; BOUREAU, Y. Lan; LECUN, Yann. Sparse Feature Learning for Deep Belief Networks. *In*: 2008, *Advances in neural information processing systems (NIPS).* p. 1185–1192.

REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali. You Only Look Once: Unified, Real-Time Object Detection. 2015.

REDMON, Joseph; FARHADI, Ali. YOLO9000: Better, Faster, Stronger. *Conference on Computer Vision and Pattern Recognition*, v. 7, n. 3, 2017.

REN, Lei; CUI, Jin; SUN, Yaqiang; CHENG, Xuejun. Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *Journal of Manufacturing Systems*, v. 43, p. 248–256, 2017.

REN, Lei; SUN, Yaqiang; CUI, Jin; ZHANG, Lin. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *Journal of Manufacturing Systems*, 2018.

ROCCO S., Claudio M.; ZIO, Enrico. A support vector machine integrated system for the classification of operation anomalies in nuclear components and systems. *Reliability Engineering and System Safety*, 2007.

ROHLMANN, Antonius; SCHMIDT, Hendrik; GAST, Ulf; KUTZNER, Ines; DAMM, Philipp; BERGMANN, Georg. In vivo measurements of the effect of whole body vibration on spinal loads. *European Spine Journal*, 2014.

ROY, R. N.; CHARBONNIER, S.; CAMPAGNE, A.; BONNET, S. Efficient mental workload estimation using task-independent EEG features. *Journal of Neural Engineering*, 2016.

SAHA, Bhaskar; GOEBEL, Kai; CHRISTOPHERSEN, Jon. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, v. 31, n. 3–4, p. 293–308, 2009.

SALEHIAN, S.; FAR, B. Embedded Real Time Blink Detection System for Driver Fatigue Monitoring. *The 27th International Conference on Software Engineering and Knowledge Engineering*, 2015.

SAMIEE, Sajjad; AZADI, Shahram; KAZEMI, Reza; NAHVI, Ali; EICHBERGER, Arno. Data fusion to develop a driver drowsiness detection system with robustness to signal loss. *Sensors (Switzerland)*, 2014.

SAN MARTIN, Gabriel; LÓPEZ DROGUETT, Enrique; MERUANE, Viviane; DAS CHAGAS MOURA, Márcio. Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring*, 2019.

SANDBERG, David; ÅKERSTEDT, Torbjörn; ANUND, Anna; KECKLUND, Göran; WAHDE, Mattias. Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. *IEEE Transactions on Intelligent Transportation Systems*, 2011.

SCHLEICHER, R.; GALLEY, N.; BRIEST, S.; GALLEY, L. Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, 2008.

SCHMIDHUBER, Jürgen. Deep Learning in neural networks: An overview. *Neural Networks*, v. 61, p. 85–117, 2015.

SI, Xiao Sheng; WANG, Wenbin; HU, Chang Hua; ZHOU, Dong Hua. Remaining useful life estimation - A review on the statistical data driven approaches. *European Journal of Operational Research*, v. 213, n. 1, p. 1–14, 2011.

SIKORSKA, J. Z.; HODKIEWICZ, M.; MA, L. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, v. 25, p. 1803–1836, 2011.

SKLET, Snorre. Safety barriers: Definition, classification, and performance. *Journal of Loss Prevention in the Process Industries*, v. 19, n. 5, p. 494–506, 2006.

SONG, Fengyi; TAN, Xiaoyang; CHEN, Songcan; ZHOU, Zhi Hua. A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognition*, v. 46, n. 12, p. 3157–3173, 2013.

SONKA, Milan; HLAVAC, Vaclav; BOYLE, Roger. *Image Processing, Analysis, and Machine Vision*. . .

SOUALHI, Abdenour; MEDJAHER, Kamal; ZERHOUNI, Noureddine. Bearing health monitoring based on hilbert-huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, v. 64, n. 1, p. 52–62, 2015.

SOUKUPOVÁ, Tereza; CECH, Jan. Real-Time Eye Blink Detection using Facial Landmarks. *In*: 2016, Ljubljana. *21st Computer Vision Winter Workshop*. Ljubljana: Slovenian Pattern Recognition Society, 2016.

STANDARDIZATION, I. O. f. *ISO 10816–7: Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts. Part 7: Rotodynamic pumps for industrial applications, including measurements on rotating shafts*. Switzerland: *[s. n.]*, 2009. .

STERN, John A.; BOYER, Donna J.; SCHROEDER, David J.; TOUCHSTONE, R. Mark; STOLIAROV, N. *Blinks, saccades, and fixation pauses during vigilance task performance: II. Gender and time-of-day*. Washington DC: *[s. n.]*, 1996.

SUTHARSSAN, Thamo; STOYANOV, Stoyan; BAILEY, Chris; ROSUNALLY, Yasmine. Prognostics and health monitoring of high power LED. *Micromachines*, v. 3, p. 78–100, 2012.

SUTRISNO, Edwin; OH, Hyunseok; VASAN, Arvind Sai Sarathi; PECHT, Michael.

Estimation of remaining useful life of ball bearings using data driven methodologies. *2012 IEEE Conference on Prognostics and Health Management*, v. 2, p. 1–7, 2012.

SZELISKI, Richard. Computer Vision : Algorithms and Applications. *Computer*, v. 5, p. 832, 2010.

TAMILSELVAN, Prasanna; WANG, Pingfeng. Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering and System Safety*, 2013.

TAYLOR, Graham W.; FERGUS, Rob; LECUN, Yann; BREGLER, Christoph. Convolutional learning of spatio-temporal features. *In*: 2010, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* p. 140–153.

TEIZER, Jochen. Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. *Advanced Engineering Informatics*, 2015.

THAMES, Lane; SCHAEFER, Dirk. Software-defined Cloud Manufacturing for Industry 4.0. *In*: 2016, *Procedia CIRP.*

TORRES, Maria E.; COLOMINAS, Marcelo A.; SCHLOTTHAUER, Gaston; FLANDRIN, Patrick. A complete ensemble empirical mode decomposition with adaptive noise. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

TRIPATHY, Abinash; AGRAWAL, Ankit; RATH, Santanu Kumar. Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 2016.

TURAGA, Pavan; CHELLAPPA, Rama; SUBRAHMANIAN, V. S.; UDREA, Octavian. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 18, n. 11, p. 1473–1488, 2008.

U.S. DEPARTMENT OF HOMELAND SECURITY. Guide for the Selection of Personal Protective Equipment for Emergency First Responders. v. I, n. November, p. 1–118, 2002. Disponível em: http://www.ncjrs.gov/pdffiles1/nij/191518.pdf

VALERA, M.; VELASTIN, S. A. Intelligent distributed surveillance systems: a review. *IEE Proceedings - Vision, Image, and Signal Processing*, v. 152, n. 2, p. 192, 2005.

VAPNIK, V. *The Nature of Statistical Learning Theory*. . v. 8 .

VERSTRAETE, David; FERRADA, Andrés; DROGUETT, Enrique López; MERUANE, Viviana; MODARRES, Mohammad. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, 2017.

VERWEY, Willem B.; ZAIDEL, David M. Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Personality and Individual Differences*, v. 28, n. 1, p. 123–142, 2000.

VIJAYAN, Vineetha; SHERLY, Elizabeth. Real time detection system of driver drowsiness based on representation learning using deep neural networks. *Journal of Intelligent and Fuzzy Systems*, v. 36, n. 3, p. 1977–1985, 2019.

VIOLA, Paul; JONES, Michael. Robust real-time object detection. *International Journal of Computer Vision*, v. 57, n. 2, p. 137–154, 2001.

VOYANT, Cyril et al. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, v. 105, p. 569–582, 2017.

VURAL, U.; AKGUL, Y. S. Eye-gaze based real-time surveillance video synopsis. *Pattern Recognition Letters*, v. 30, n. 12, p. 1151–1159, 2009.

WANG, Ching Wei; HUNTER, Andrew; GRAVILL, Neil; MATUSIEWICZ, Simon. Real time pose recognition of covered human for diagnosis of sleep apnoea. *Computerized Medical Imaging and Graphics*, v. 34, n. 6, p. 523–533, 2010.

WANG, Lipo. Support Vector Machines : Theory and Applications. 2005.

WANG, Mei; GUO, Lin; CHEN, Wen Yuan. Blink detection using Adaboost and contour circle for fatigue recognition. *Computers and Electrical Engineering*, v. 58, p. 502–512, 2017.

WANG, Shiyong; WAN, Jiafu; ZHANG, Daqiang; LI, Di; ZHANG, Chunhua. Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination. *Computer Networks*, 2016.

WIDODO, Achmad; YANG, Bo Suk. Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications*, v. 38, n. 7, p. 8430–8437, 2011.

WILSON, Marian; PERMITO, Regan; ENGLISH, Ashley; ALBRITTON, Sandra; COOGLE, Carlana; VAN DONGEN, Hans P. A. Performance and sleepiness in nurses working 12-h day shifts or night shifts in a community hospital. *Accident Analysis and Prevention*, 2019.

WRIGHT, Stephen J. *Primal-Dual Interior-Point Methods*. . .

WU, Zhaohua; HUANG, Norden E. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Anal*, v. 2009, n. 1, p. 1–41, 2009.

YAN, Ruqiang; GAO, Robert X.; CHEN, Xuefeng. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, v. 96, n. PART A, p. 1–15, 2014.

YANG, Yu; YU, Dejie; CHENG, Junsheng. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement: Journal of the International Measurement Confederation*, v. 40, n. 9–10, p. 943–950, 2007.

YILMAZ, Alper; JAVED, Omar; SHAH, Mubarak. Object tracking: A survey. *ACM Comput.*

Surv., v. 38, n. 4, p. 13, 2006.

ZAFEIRIOU, Stefanos; ZHANG, Cha; ZHANG, Zhengyou. A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, 2015.

ZELINSKY, Gregory J.; PENG, Y.; SAMARAS, D. Eye can read your mind: decoding gaze fixations to reveal categorical search targets. *Journal of Vision*, v. 13, n. 14, p. 1–13, 2013.

ZHANG, Chao; WU, Xiaopei; ZHANG, Lei; HE, Xuan; LV, Zhao. Simultaneous detection of blink and heart rate using multi-channel ICA from smart phone videos. *Biomedical Signal Processing and Control*, v. 33, p. 189–200, 2017.

ZHANG, Ting et al. Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *International Journal of Psychophysiology*, 2015.

ZHAO, Qijie; YUAN, Xinming; TU, Dawei; LU, Jianxia. Eye moving behaviors identification for gaze tracking interaction. *Journal on Multimodal User Interfaces*, 2015.