



**PARKINSON PATTERNS QUANTIFICATION MODELED AS ABNORMAL
EVENTS USING A GENERATIVE DEEP LEARNING STRATEGY**

EDGAR YESID RANGEL PIESCHACON

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2023

**PARKINSON PATTERNS QUANTIFICATION MODELED AS ABNORMAL
EVENTS USING A GENERATIVE DEEP LEARNING STRATEGY**

EDGAR YESID RANGEL PIESCHACON

**Research work in partial fulfillment of the requirements for the degree of:
Magíster en Ingeniería de Sistemas e Informática**

Advisor:

Fabio Martínez Carrillo

Ph.D in Systems and Computer Engineering

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2023

ACKNOWLEDGEMENTS

The author expresses his acknowledgement:

To my family, Yesid Rangel and Sandra Pieschacon for being a really great parents who teach me many things, to my sister and niece Pilar Rangel and Juan David Rivera. A great thanks to Professor Fabio Martinez for being a great academic and professional guide, but also, for being a friend and a support in all this process. In general, to the research group Biomedical Imaging, Vision and Learning Laboratory for supporting most of the technical and theoretical components needed. Thanks also to all the colleagues who in some way made possible the achievements that today are consolidated in this document.

Special thanks to all the special people in my life, family members, and relatives who have been unconditionally and important during this process and deserve to receive all the credits.

Finally, thanks to the Escuela de Ingeniería de Sistemas e Informática (EISI) and Universidad Industrial de Santander (UIS), and other staff who made this process possible.

CONTENTS

	page
INTRODUCTION	11
1. FUNDAMENTALS AND PREVIOUS WORK	14
1.1. Learning algorithms	14
1.1.1 Discriminatory Learning	15
1.1.2 Adversarial-Generative Learning	16
1.2. Learning Parkinsonian gait representations	17
1.2.1 Anomaly detection framework	19
2. Research Problem	21
3. OBJECTIVES	23
4. PROPOSED APPROACH	24
4.1. A volumetric autoencoder to recover gait embedding patterns	24
4.2. Auxiliary task to discriminate reconstructions	26
4.3. A Digital gait biomarker from anomaly embeddings	27
4.4. Statistical validation setup	28
4.4.1 Variance analysis from Homoscedasticity	28
4.4.2 Shapeness analysis from ChiSquare	29
5. EXPERIMENTAL SETUP	32
5.1. Datasets	32
5.1.1 Captured dataset	32
5.1.2 External dataset validation	33

5.2. Model configuration 35

6. EVALUATION AND RESULTS 36

6.1. Captured dataset results 36

6.2. Statistical validation of captured dataset 37

6.3. External dataset results 37

6.4. Statistical validation of external dataset 39

6.5. Parkinson training validation 39

7. DISCUSSION 41

8. CONCLUSIONS AND FUTURE WORK 44

BIBLIOGRAPHY 45

APPENDICES 50

LIST OF FIGURES

	page
Figure 1. Deep learning schemes generalized	15
Figure 2. Pipeline of the proposed work to model and separate between classes	25
Figure 3. Standard metrics limitation for unseen samples	27
Figure 4. Capture setup used for recording	32
Figure 5. Examples of the recorded dataset	33
Figure 6. Examples of external dataset	34
Figure 7. Data distribution of captured dataset	38
Figure 8. Data distribution of external dataset	40

LIST OF TABLES

	page
Table 1. Generator and Discriminator Networks structure summary	35
Table 2. Model performance in captured dataset	36
Table 3. Model statistical validation of captured dataset	37
Table 4. Model performance in external dataset	38
Table 5. Statistical validation of external dataset	39
Table 6. Model performance by training with parkinson samples	40

LIST OF APPENDICES

	page
Appendix A. Academic Products	50

ABSTRACT

TITLE: PARKINSON PATTERNS QUANTIFICATION MODELED AS ABNORMAL EVENTS USING A GENERATIVE DEEP LEARNING STRATEGY *

AUTHOR: EDGAR YESID RANGEL PIESCHACON **

KEYWORDS: ANOMALY DETECTION, DEEP LEARNING, PARKINSON GAIT, GENERATIVE REPRESENTATIONS, EMBEDDED REPRESENTATIONS, PARKINSON DISEASE.

DESCRIPTION: Parkinson's Disease (PD) is the world's second most common neurodegenerative disorder, affecting more than 6.2 million people. This disease is associated with the deficiency of the neurotransmitter dopamine, associated with the control of voluntary movements. In consequence, PD is associated with gait movement disorders, such as bradykinesia, stiffness, tremors and postural instability, caused by progressive dopamine deficiency. Today, some approaches have implemented learning representations to quantify kinematic patterns during locomotion, supporting clinical procedures such as diagnosis and treatment planning. These approaches assumes a large amount of stratified and labeled data to optimize discriminative representations. Nonetheless, these considerations may restrict the approaches to be operable in real scenarios during clinical practice. This work introduces a self-supervised generative representation to learn gait-motion-related patterns, under the pretext of video reconstruction and an anomaly detection framework. This architecture is trained following a one-class weakly supervised learning to avoid inter-class variance and approach the multiple relationships that represent locomotion. The proposed approach was validated with two datasets containing a total of 133 patients with control, parkinson and knee-osteoarthritis diagnosis, achieving AUC of 80.6% and 75%, homocedasticity level of 80% and 66.7%; and shapeness level of 70% and 66.7% respectively. These results evaluate the model in its discrimination task considering its generalization for unseen samples.

* Research work

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Fabio Martínez Carrillo, Ph.D.

RESUMEN

TÍTULO: CUANTIFICACIÓN DE PATRONES DE PARKINSON MODELADOS COMO EVENTOS ANORMALES UTILIZANDO UNA ESTRATEGIA DE APRENDIZAJE PROFUNDO GENERATIVO *

AUTOR: EDGAR YESID RANGEL PIESCHACON **

PALABRAS CLAVE: DETECCIÓN ANOMALA, APRENDIZAJE PROFUNDO, MARCHA PARKINSONIANA, REPRESENTACIONES GENERATIVAS, REPRESENTACIONES EMBEBIDAS, ENFERMEDAD DEL PARKINSON

DESCRIPCIÓN: La enfermedad de Parkinson (EP) es el segundo trastorno neurodegenerativo más frecuente del mundo y afecta a más de 6,2 millones de personas. Esta enfermedad se asocia a la deficiencia del neurotransmisor dopamina, relacionado con el control de los movimientos voluntarios. En consecuencia, la EP se asocia a trastornos del movimiento de la marcha, como bradicinesia, rigidez, temblores e inestabilidad postural, causados por la deficiencia progresiva de dopamina. En la actualidad, algunos enfoques han implementado representaciones de aprendizaje para cuantificar los patrones cinemáticos durante la locomoción, apoyando procedimientos clínicos como el diagnóstico y la planificación del tratamiento. Estos enfoques asumen una gran cantidad de datos estratificados y etiquetados para optimizar las representaciones discriminativas. Sin embargo, estas consideraciones pueden restringir los enfoques para ser operables en escenarios reales durante la práctica clínica. Este trabajo introduce una representación generativa autosupervisada para aprender patrones relacionados con el movimiento de la marcha, con el pretexto de la reconstrucción de vídeo y un marco de detección de anomalías. Esta arquitectura se entrena siguiendo un aprendizaje débilmente supervisado de una clase para evitar la varianza interclase y aproximarse a las múltiples relaciones que representan la locomoción. El enfoque propuesto se validó con dos conjuntos de datos que contenían un total de 133 pacientes con diagnóstico de control, parkinson y artrosis de rodilla, alcanzando AUC de 80,6% y 75%, nivel de homocedasticidad de 80% y 66,7%; y nivel de conformación de 70% y 66,7% respectivamente. Estos resultados evalúan el modelo en su tarea de discriminación considerando su generalización para muestras no vistas.

* Trabajo de investigación

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.

INTRODUCTION

Parkinson’s Disease (PD) is the world’s second most common neurodegenerative disorder, affecting more than 6.2 million people ^{1 2}. According to the World Health Organization, this number will increase by more than 12 million by 2030 ³. PD is characterized by the progressive loss of dopamine, a neurotransmitter involved in the execution of voluntary movements. For this reason, the main diagnostic support is based on the observation and analysis of progressive motor disorders, such as tremor, rigidity, slowness of movement (bradykinesia), postural instability, among many other related symptoms ⁴. Despite significant improvements in determining the causes of the disease and its various symptoms, today, there is not a definitive and universal biomarker to characterize, diagnose, and follow the patient progression of PD patients.

Particularly, the gait is a multi-factorial and complex locomotion process that involves several subsystems. The associated kinematics patterns are typically recovered over standard marker-based setups, that coarsely approximate complex motion behaviors, resulting restrictive, intrusive and, altering natural postural gestures for PD description. Alternative, markerless video strategies together with discriminative learning approximations have emerged as key solutions

¹ Theo Vos et al. “Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”. In: *The Lancet* 390.10100 (2017), pp. 1211–1259.

² E Ray Dorsey and Bastiaan R Bloem. “The Parkinson pandemic—a call to action”. In: *JAMA neurology* 75.1 (2018), pp. 9–10.

³ World Health Organization. *Neurological disorders: public health challenges*. World Health Organization, 2006.

⁴ R Balestrino and AHV Schapira. “Parkinson disease”. In: *European journal of neurology* 27.1 (2020), pp. 27–42.

to support the PD characterization and classification from other diseases^{5 6 7 8 9}. These methodologies have been successful in controlled studies but strongly require a stratified, balanced, and well-labeled dataset to avoid overfitting. Besides, these approaches are biased to the physician’s experience to determine the disease and limit the quantification to general scale indexes^{10 11}. Even worse, these approaches solve classification tasks but remain limited on further explanation about data representation to define the generalization capability w.r.t new patients.

This work introduces a deep generative and anomaly architecture to learn a hidden descriptor to represent locomotion patterns. Following a weakly supervised methodology, a 3D convolutional network is self-trained under a gait video reconstruction pretext task. Then, the resultant embedding representation encodes complex dynamic gait relationships, captured from control population, that allow to discriminate parkinson patients. The main contributions of this work are summarized as follows:

-
- ⁵ Navleen Kour, Sakshi Arora, et al. “Computer-vision based diagnosis of Parkinson’s disease via gait: a survey”. In: *IEEE Access* 7 (2019), pp. 156620–156645.
 - ⁶ Luis Carlos Guayacán, Edgar Rangel, and Fabio Martínez. “Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 3688–3691.
 - ⁷ Renfei Sun et al. “Convolutional 3D attention network for video based freezing of gait recognition”. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE. 2018, pp. 1–7.
 - ⁸ Luis C Guayacán and Fabio Martínez. “Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network”. In: *Journal of biomedical informatics* 123 (2021), p. 103935.
 - ⁹ Michael H Li et al. “Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation”. In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), pp. 1–13.
 - ¹⁰ Geert Litjens et al. “A survey on deep learning in medical image analysis”. In: *Medical image analysis* 42 (2017), pp. 60–88.
 - ¹¹ Christopher G Goetz et al. “Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results”. In: *Movement disorders: official journal of the Movement Disorder Society* 23.15 (2008), pp. 2129–2170.

- A new digital biomarker coded as an embedding vector with the capability to represent hidden kinematic relationships of Parkinson disease.
- A 3D Convolutional GAN net dedicated to learn spatio-temporal patterns of gait video-sequences. This architecture integrates an auto-encoder net to learn video patterns in reconstruction tasks and a complementary decoder that discriminates among reconstructed and original video sequences.
- A statistical test framework to validate the capability of the approach in terms of generalization, coverage of data, and discrimination capability for any class with different groups between them, *i.e.* evaluate the generalization of Parkinsonian patients, at different stages of the disease, with respect to a control population.

1. FUNDAMENTALS AND PREVIOUS WORK

The present research proposal addresses topics related to the automatic modeling of spatio-temporal patterns related to PD. Hence, in this section will be defined concepts related to machine learning schemes, video analysis as well as approximations to the modeling on video of PD.

1.1. Learning algorithms

Machine learning methods namely use a set of labeled information to minimize a model representation (supervised) and perform a task related to classification and representation. Un-supervised learning, in contrast, tries to recover patterns without the use of any high-level information. In figure 1 is illustrated both learning schemes. Nowadays the most common methodology in both learning schemes, are deep neural networks which define a set of non-linear units (neurons), grouped at different levels of processing (layers) to minimize some objective function. Specifically, the units comprise an activation a and parameters $\theta = \{\mathbf{W}, \mathbf{B}\}$, where \mathbf{W} is a set of weights and \mathbf{B} is the bias. The output at each unit is then described as $a = \sigma(\mathbf{W}\Delta\mathbf{x} + \mathbf{b})$. The hierarchical composition of units is then modelled as $f(\mathbf{x}; \theta) = \sigma(\mathbf{w}\Delta\sigma(\mathbf{w}\Delta\dots\sigma(\mathbf{w}\Delta\mathbf{x} + b)) + b)$ ¹⁰. In such model, θ and the unit arrangements constitutes the learning representation of a particular trained problem. Moreover the model $f(\mathbf{x}; \theta)$ can be operated with convolutions ($*$), linear combination (\cdot), normalization, among others operations (Δ). Every w and b will vary depending of the operation adjusted and generate the output $f(\mathbf{x}; \theta) = \hat{y}$. For instance, the output can represent a probability on a classification problem or a newly generated sample \hat{x} . According to the rule implemented during training, the approaches can be grouped as discriminatory and generative models. In next subsection will expand a description of these concepts.

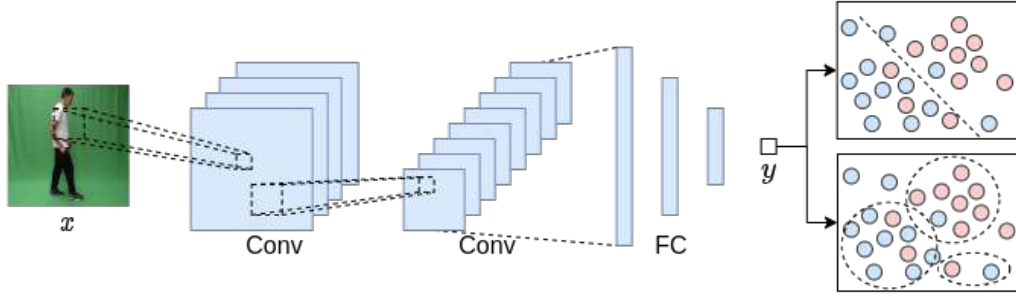


Figure 1. Learning schemes generalized, where supervised learning focus on separating among classes and unsupervised learning focus on grouping among classes.

1.1.1. Discriminatory Learning The discriminatory task is generally performed into a supervised learning scheme, following a loss function $L(y, \hat{y})$ that tries to find the best parameters θ given the minimization between a set of high-level labels y and the corresponding estimations \hat{y} obtained by feeding a data point input \mathbf{x} . The model $f(\mathbf{x}; \theta)$ commonly represents the probability of \mathbf{x} being \hat{y}_i ($P(\hat{y}^{(i)}|\mathbf{x}; \theta)$) where (i) correspond to a single class on the dataset in the final activation layer (generally a softmax function). In deep learning, the most common loss function L , into supervised schemes is the cross-entropy (equation 1) ¹⁰ where it sums the logarithms probabilities of the data points \mathbf{x} multiplied by the associated probability of the label. This expression is given by:

$$L(y, \hat{y}^{(i)}) = -P(y^{(i)}|\mathbf{x}; \theta) \sum_{n=1}^N \log [P(\hat{y}^{(i)}|\mathbf{x}; \theta)] \quad (1)$$

The main limitation of discriminatory schemes is the requirement of high-volume semantic annotations given or requiring human supervision. These annotations may be biased by experts, and for some applications are unbalanced, which could be biased from the proper separation among data classes. Also, the loss function in equation 1 has limitations presented in the logarithm function, where $P(\hat{y}^{(i)}|\mathbf{x}; \theta)$ is nearly to 0, the loss can increase to infinity generating

numerical instability and computational cost to minimize the function ¹².

1.1.2. Adversarial-Generative Learning Alternatively, the generative task, as part of unsupervised schemes, tries to generate or reconstruct samples from training data, learning directly from input features, and forcing the representation to cover the whole spectrum that represents the problem. In general, the generative task follows a minimization rule as $\|\mathbf{x} - \hat{\mathbf{x}}\|$, where $\hat{\mathbf{x}}$ is the obtained via reconstruction, and tries to be closer to the input represented in features \mathbf{x} . The current deep learning approaches use two different hierarchical schemes to represent generative problems: the auto encoders and the generative adversarial networks (GANs) ¹³. While the auto encoders try to learn embedding vectors, namely conditioned by specific distributions, the GANs try to learn a flexible latent subspace without parametric distribution conditions ¹⁰. Hence, GANs can discover new hidden patterns, embedded into a compact space, that may be strongly associated with the input.

The GANs introduced a new training adversarial scheme of a couple of networks: the generator and discriminator ¹⁴. On the one hand, the generator may start with a random vector that is fully correlated from net layers until it achieves the original dimension of interest data. On the other hand, the discriminator receives the generated data and a real sample and tries to classify it as false (generated) input or the corresponding from a real dataset. In this adversarial game, the generated network is forced each time to obtain better-generated images to cheat the discriminator network. During this process, a very relevant thing is the learning of a latent space vector that compactly represents multidimensional data into compact embedding vectors. The optimization rule of GANs is described in equation 2 as follows:

¹² Varun Chandola, Arindam Banerjee, and Vipin Kumar. “Anomaly detection: A survey”. In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58.

¹³ Raghavendra Chalapathy and Sanjay Chawla. “Deep learning for anomaly detection: A survey”. In: *arXiv preprint arXiv:1901.03407* (2019).

¹⁴ Ian J Goodfellow et al. “Generative adversarial networks”. In: *arXiv preprint arXiv:1406.2661* (2014).

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(z)} [\log 1 - D(G(\mathbf{z}))] \quad (2)$$

In this equation, the \mathbf{x} is the input real data, \mathbf{z} is a random input vector associated with a latent space. Hence, the expression $\mathbf{x} \sim p_{data}(\mathbf{x})$ describes the approximation of x to the probability distribution function (pdf) of data in the point x , $\mathbf{z} \sim p_z(z)$ means the approximation of z to pdf of latent vectors in the point z . Also, $G(\cdot)$ is the generation of a new sample of \cdot input, $D(\cdot)$ is the classification of \cdot of being real and \mathbb{E} is the arithmetic mean. An advantage of generative learning comes in the loss function used for the models, *e.g.* equation 3 shows a subtraction of the input x and generated data \hat{x} to determine the loss value in generator (G).

$$L_G(\mathbf{x}, \hat{\mathbf{x}}) = |\mathbf{x} - \hat{\mathbf{x}}|_2 \quad (3)$$

Generally, for unsupervised learning schemes, the loss functions follow a subtraction operation which makes them more numerically stable and the same or less computational cost of supervised schemes. However, is challenging to learn commonalities within data in a complex and high dimensional space ¹³.

1.2. Learning Parkinsonian gait representations

Deep discriminative learning is nowadays the standard methodology in computer vision challenges, demonstrating remarkable results in very different domains. For instance, the Parkinson characterization is achieved from sensor-based and vision-based approaches, following a supervised scheme to capture main observed relationships and to generate a particular prediction about condition of the patients ⁵. These approaches in general are dedicated to classify and discriminate between a control population and patients with the Parkinson condition. The sensor based approaches capture kinematics from motion signals, approximating to PD classification, but in much of the cases results marker-invasive, alter natural gestures, and only have recog-

dition capabilities in advanced stages of the disease ¹⁵. Contrary, the vision-based approaches exploit postural and dynamic features, from video recordings, but the representations underlies on supervised schemes that requires a large amount of labeled data to learn the inter and intra variability among classes ^{6 7 8 9}. Also, these learning methodologies require that training data have well-balanced conditions among classes, *i.e.*, to have the same proportion of sample observations for each of the considered class ¹³.

Unsupervised, semi-supervised and weakly supervised approaches have emerged as a key alternative to model biomedical problems, with significative variability among observations but limited training samples. However, to the best of our knowledge, these learning methods have been poorly explored and exploited in Parkinson characterization, with some preliminar alternatives that use principles of Minimum Distance Classifiers and K-means Clustering ^{5 16 17 18 19 20}. In such sense, the PD modelling from non-supervised perspective may be addressed from reconstruction, prediction and generative tasks ²¹, that help to determine samples distribu-

-
- ¹⁵ K Sugandhi, Farha Fatina Wahid, and G Raju. “Feature extraction methods for human gait recognition—a survey”. In: *International Conference on Advances in Computing and Data Sciences*. Springer. 2016, pp. 377–385.
- ¹⁶ Lars Schmarje et al. “A survey on semi-, self-and unsupervised learning for image classification”. In: *IEEE Access* 9 (2021), pp. 82146–82168.
- ¹⁷ Chien-Wen Cho et al. “A vision-based analysis system for gait recognition in patients with Parkinson’s disease”. In: *Expert Systems with applications* 36.3 (2009), pp. 7033–7039.
- ¹⁸ Shih-Wei Chen et al. “Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis”. In: *Biomedical engineering online* 10.1 (2011), pp. 1–21.
- ¹⁹ Sven Nõmm et al. “An alternative approach to distinguish movements of parkinson disease patients”. In: *IFAC-PapersOnLine* 49.19 (2016), pp. 272–276.
- ²⁰ Sara Soltaninejad et al. “Body movement monitoring for Parkinson’s disease patients using a smart sensor based non-invasive technique”. In: *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE. 2018, pp. 1–6.
- ²¹ B Ravi Kiran, Dilip Mathew Thomas, and Ranjith Parakkal. “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos”. In: *Journal of Imaging* 4.2 (2018), p. 36.

tion and determine future postural and kinematic events. In fact, the PD pattern distribution results key to understand multi-factorial nature of PD, being determinant to define variations such as laterality affectation of disease, abnormality sources, but also to define patient prognosis, emulating the develop of a particular patient during the gait.

1.2.1. Anomaly detection framework Parkinson modelling may be addressed from one class learning (a.k.a anomaly detection or abnormal learning), that can be used to learn relevant characteristics from the data ¹³. In such sense, in the literature have been proposed different strategies that model anomaly detection from low level image information (background, color, textural patterns) or from semantically object descriptions (motion information, object shape) ²². For instance, Ravanbakhsh *et al.* ²³ and Schlegl *et al.* ²⁴ used a Generative Adversarial Networks (GANs), to output synthetic images and adjust the discriminator net. Then, the adjusted discriminator is used to determine anomaly images that are outliers from adjusted distribution ²⁵. Schlegl *et al.* proposes later a generative anomaly net that fix input distribution but using input images rather than random input vectors, and showing a better performance

²² Rashmiranjan Nayak, Umesh Chandra Pati, and Santos Kumar Das. “A comprehensive review on deep learning-based methods for video anomaly detection”. In: *Image and Vision Computing* 106 (2021), p. 104078.

²³ Mahdyar Ravanbakhsh et al. “Abnormal event detection in videos using generative adversarial nets”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 1577–1581.

²⁴ Thomas Schlegl et al. “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery”. In: *International conference on information processing in medical imaging*. Springer. 2017, pp. 146–157.

²⁵ Federico Di Mattia et al. “A survey on gans for anomaly detection”. In: *arXiv preprint arXiv:1906.11632* (2019).

in the detection task ²⁶. Also, Akcay *et al.* ²⁷ ²⁸ proposes the GANomaly, an adversarial autoencoder, which uses the image as input instead of a random variable Z associated to the selected input distribution. This net generates a latent space that better represent trained images and finally, Zenati *et al.* proposes a BiGAN model, that consist on a vanilla GAN with an Encoder to learn the latent space from images and reconstruct the original image using the generate embedding vector ²⁹.

Despite of remarkable approximations of anomaly detection, these approach lack of explainability for the models that describe output probabilities from observation inputs ²². To cover such limitations some works have carried out ablation studies to determine main components that generalize anomaly tasks ³⁰ ³¹. These approaches however provide poor explanations about learned distribution.

-
- ²⁶ Thomas Schlegl *et al.* “f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks”. In: *Medical image analysis* 54 (2019), pp. 30–44.
- ²⁷ Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. “Ganomaly: Semi-supervised anomaly detection via adversarial training”. In: *Asian conference on computer vision*. Springer. 2018, pp. 622–637.
- ²⁸ Samet Akçay, Amir Atapour-Abarghouei, and Toby P Breckon. “Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection”. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019, pp. 1–8.
- ²⁹ Houssam Zenati *et al.* “Efficient gan-based anomaly detection”. In: *arXiv preprint arXiv:1802.06222* (2018).
- ³⁰ Mahdyar Ravanbakhsh *et al.* “Training adversarial discriminators for cross-channel abnormal event detection in crowds”. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2019, pp. 1896–1904.
- ³¹ Xu Han, Xiaohui Chen, and Li-Ping Liu. “GAN Ensemble for Anomaly Detection”. In: *arXiv preprint arXiv:2012.07988* (2020).

2. Research Problem

Main PD affectations are associated with locomotion disorders due to dopamine deficiency. During gait, the PD locomotion is characterized by slowness of movement (bradykinesia), postural instability, body stiffness, reduced step length, among others. Despite main motion PD features, nowadays there is not a definitive biomarker to analyze and quantify disease progression. Hence, the motor stage and disease progression are coarsely grouped from scales that integrate several symptoms and movement disorder observations. These observations are however subjective with evidence in the literature of false-positive PD diagnosis of 35% at the initial diagnosis and 24% at final diagnosis ³².

The kinematic gait analysis quantitatively supports motor pattern quantification but their implementation is restricted to follow a set of markers attached to the body during the locomotion ¹⁵. These methods are invasive, alter the natural gesture and greatly simplify the dynamics of locomotion to a reduced set of trajectories recorded throughout the gait cycle. Markerless strategies have been proposed to complement classical gait pattern analysis, as well as, to discover new relationships between the spatio-temporal patterns modeled in video sequences and the PD progression. Following this line, the machine and deep learning approaches have been implemented to resolve tasks related to the identification and classification of PD patients w.r.t other patient populations. The main issue with these typical learning schemes is the discriminative character and the explicit adjusting of representation concerning annotated data, which has an inherent subjectivity and coarse representation of the disease. In consequence, sometimes these models focus on non-motor instances, background, or certain morphological characteristics of a particular dataset, losing the principal objective of the training scheme. Besides, such classical learning schemes namely require a balanced set of training, which results unnatural

³² HEIDI SPLETE. “Avoid Diagnostic Pitfalls For Parkinson’s Disease”. In: *Internal Medicine News* 39.17 (2006), pp. 31–31.

in a daily clinical routine scenario. From such principal described points, we propose the next research question:

Research Question

How much unsupervised generative learning representation, into an anomaly detection scheme, can contribute to the quantification of PD patterns?

3. OBJECTIVES

General Objective

- To propose a deep generative strategy, under an anomaly detection scheme, to quantify parkinsonian gait spatial-temporal patterns.

Specific Objectives

- To select a Parkinson's gait dataset that includes Parkinson patients and control subjects.
- To define a deep generative scheme to learn motor patterns related to the parkinsonian gait.
- To develop an anomaly training scheme to capture embedding latent vectors with full correspondence with a particular instance of the dataset.
- To validate the proposed methodology regarding the capability of capture abnormal pattern associated with PD from learned deep vectors.

4. PROPOSED APPROACH

This work introduces a digital PD biomarker that embedded gait motor patterns, from anomaly video reconstruction task. Contrary to typical classification modeling, we are dedicated to deal with one class learning, *i.e.*, only to learn control gait patterns, approaching the high variability on training samples, without using explicit disease labels. Hence, we hypothesize that a digital biomarker of the disease can be modeled as a mixture of distributions, composed of samples that were labeled as outliers, from learned representation. In consequence, we analyze the embedding, reconstruction, and discrimination space to later define rules to separate Parkinson from control vectors, during test validation. The general pipeline of the proposed approach is illustrated in Figure 2.

4.1. A volumetric autoencoder to recover gait embedding patterns

Here, we are interested on capture complex dynamic interactions during locomotion, observed in videos as spatio-temporal textural interactions. From a self supervised strategy (video-reconstruction task), we implemented a 3D deep autoencoder that project videos into low-dimensional vectors, learning the complex gait dynamics into a latent space (see the architecture in Figure 2-a). For doing so, 3D convolutional blocks were implemented, structured hierarchically, with main purpose to carried out a spatio-temporal reduction while increase feature descriptions. Formally, a gait sequence $\mathbf{x} \in \mathbb{N}^{f \times h \times w \times c}$, where f denotes the number of temporal frames, $(h \times w)$ are the spatial dimensions, and c is the number of color channels in the video. This sequence is received as input in the convolutional block which is convolved with a kernel κ of dimensions (k_t, k_h, k_w) , where k_t convolves on the temporal axis and k_h, k_w on the spatial axes. At each level l of processing, we obtain a new volume $\mathbf{x}_l \in \mathbb{Z}^{f/2^l \times h/2^l \times w/2^l \times 2^l c}$ that represents a bank of spatio-temporal feature maps. Each of these volumetric features are dedicated to stand out relevant gait patterns in a \mathbf{z}_G reduced projection, that summarize a

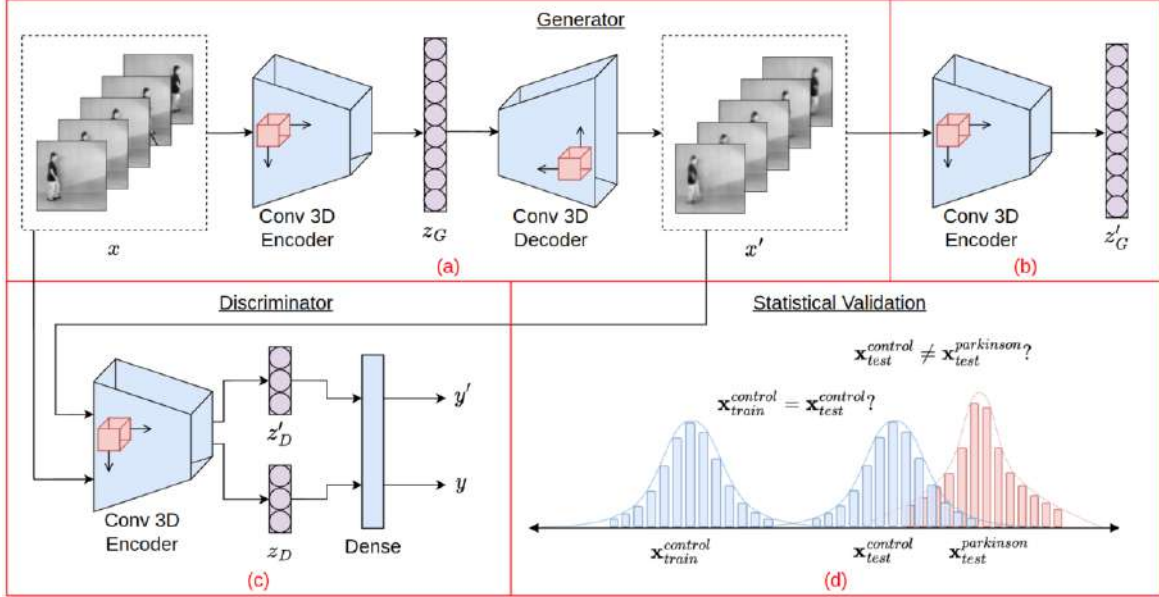


Figure 2. Pipeline of the proposed model separated in (a) volumetric auto-encoder to recover gait patterns, (b) Digital gait biomarker, (c) Auxiliary task to discriminate reconstructions, and (d) statistical validation of learned classes distributions

multiscale gait motion representation.

The resultant embedding vector \mathbf{z}_G encode principal dynamic non-linear correlations, which are necessary to achieve a video reconstruction \mathbf{x}' . In this study, the validated datasets are recorded from a relative static background, so, the major dependencies to achieve an effective reconstruction lies in temporal and dynamic information expressed during the gait. Here, we adopt \mathbf{z}_G as a digital gait biomarker that, among others, allows to study motion abnormalities associated to the Parkinson disease.

To complete end-to-end learning, 3D transposed convolutional blocks were implemented as decoder, positioned in a symmetrical configuration regarding the encoder levels, and upsampling spatio-temporal dimensions to recover original video-sequence. Formally, having the embedded feature vector $\mathbf{z}_G \in \mathbb{Z}^n$ with n coded features, we obtain $\mathbf{x}'_l \in \mathbb{Z}^{2^l f \times 2^l h \times 2^l w \times c / 2^l}$ volumes from transpose convolutional blocks until obtaining a video reconstruction $\mathbf{x}' \in \mathbb{N}^{f \times h \times w \times c}$. The quality of reconstruction is key to guarantee the deep representation learning in the autoencoder part of generator. To do this, an L_1 loss is implemented between \mathbf{x} and \mathbf{x}' and its named

contextual loss: $L_{con} = \|\mathbf{x} - \mathbf{x}'\|_1$.

4.2. Auxiliary task to discriminate reconstructions

From a generative learning, the capability of the deep representations to code locomotion patterns may be expressed in the quality of video reconstructions \mathbf{x}' . Hence, we hypothesize that embedding descriptors \mathbf{z}_G that properly reproduce videos \mathbf{x}' should encode sufficient kinematic information of trained class, allowing to discriminate among locomotion populations, *i.e.* between control and Parkinson samples.

To measure this reconstruction capability, an auxiliary task is here introduced to receive tuples with original and reconstructed videos $(\mathbf{x}, \mathbf{x}')$, and output a discriminatory decision $\mathbf{y} = \{y, y'\}$, regarding video source. In such case, y corresponds to the label for real videos, while y' as labels for embeddings from reconstructed sequences. For doing so, we implement an adversarial L_2 loss, expressed as: $L_{adv} = \|\mathbf{z}_D - \mathbf{z}'_D\|_2$. In such case, for large differences between $(\mathbf{z}_D, \mathbf{z}'_D)$ it will be a significant error that will be propagated to the generator. It should be noted that such minimization rule optimize only the generator. Then discriminator is only minimized following a classical equally weighted cross-entropy rule, as: $L_{disc} = \frac{\log(y) + \log(1-y')}{2}$.

The auxiliary task to monitor video reconstruction is implemented from a discriminatory convolutional net that follows the same structure that encoder in Figure 2-a, which halve the spatio-temporal dimension while increase the features and finally dense layer determine its realness level (see in Figure 2-c.). Interestingly, from such deep convolutional representation the input videos are projected to an embedding vector $\mathbf{z}_D \in \mathbb{Z}^m$ with m coded features, which thereafter may be used as latent vectors descriptors that also encode motion and realness information. To guarantee an optimal coding into low-dimensional embeddings, the reconstructed video \mathbf{x}' is mapped to an additional encoder projecting representation basis in a \mathbf{z}'_G embedding. In such sense, \mathbf{z}_G and \mathbf{z}'_G must be similar, and lead to \mathbf{x} and \mathbf{x}' to be equal which helps in generalization of the generator, following an encoder L_2 loss: $L_{enc} = \|\mathbf{z}_G - \mathbf{z}'_G\|_2$.

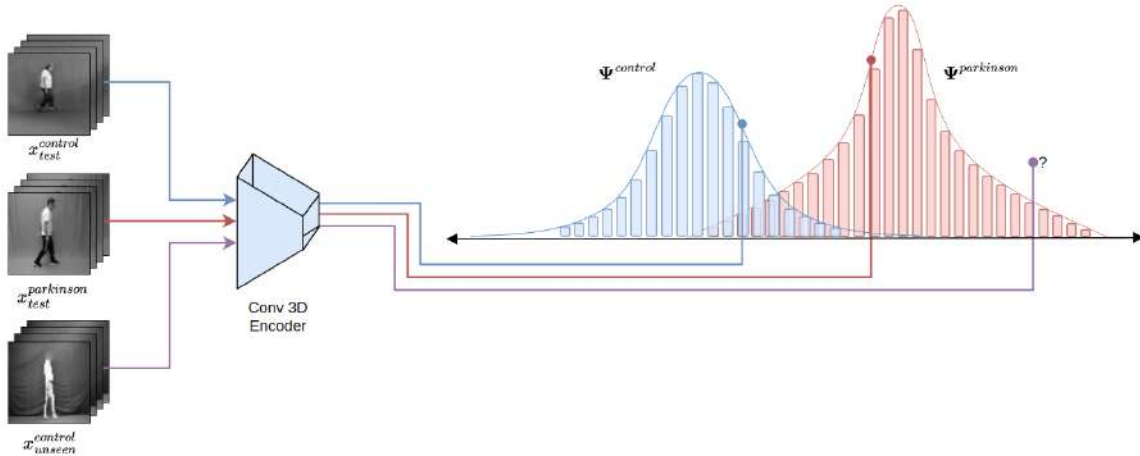


Figure 3. Field of action of standard metrics of the model, where the dataset used only cover the intersection area but the model performance for new samples is not being evaluated

4.3. A Digital gait biomarker from anomaly embeddings

The video samples are high-dimensional motor observations that can be projected into a low-dimensional embedding space, through the proposed model. Formally, each video sample is an independent and random variable $\mathbf{x}_\ell^{(i)}$ from the class (i) that follows a distribution $\mathbf{x}_\ell^{(i)} \in \Psi^{(i)}[\mu(\mathbf{x}^{(i)}), \sigma(\mathbf{x}^{(i)})]$ with mean $\mu(\mathbf{x}^{(i)})$, and standard deviation $\sigma(\mathbf{x}^{(i)})$. We then considered the proposed model as an operator that transform each sample $F(\mathbf{x}_\ell^{(i)})$ into a low dimensional space, while preserves the original distribution, as: $F(\mathbf{x}_\ell^{(i)}) \in \Psi^{(i)}[F(\mu(\mathbf{x}^{(i)})), F(\sigma(\mathbf{x}^{(i)}))]$. From this assumption we can measure statistical properties over low-dimensional space and explore properties as the generalization of the modelling.

Hence, we can adopt a new digital kinematic descriptor by considering embedding vector differences between $(\mathbf{z}_G, \mathbf{z}'_G)$. For instance, large difference between $\mathbf{z}_G, \mathbf{z}'_G$ may suggest a new motion class, regarding the original distribution of training. From such approximation, we can model an scheme of one-class learning (in this case, anomaly learning) over the video distributions from the low-embedding differences observations. This scheme learns data distribution without any label constraint. Furthermore, if we train the architecture only with videos of a control population (c), we can define a discriminatory problem from the reconstruction, by

inducing: $\|\mathbf{z}_G - \mathbf{z}'_G\|_2 \leq \tau \rightarrow \mathbf{c} \wedge \|\mathbf{z}_G - \mathbf{z}'_G\|_2 > \tau \rightarrow \mathbf{p}$, where \mathbf{p} is a label imposed to a video with a significant error reconstruction and projected to a Parkinson population.

4.4. Statistical validation setup

This new discriminatory descriptor can be validated following standard metrics into binary projection $\hat{y} = \{\mathbf{c}, \mathbf{p}\}$. For a particular threshold τ we can recover metrics such as the accuracy, the precision and the recall. Also, ROC-AUC (the Area Under the Curve) can estimate a performance by iterating over different τ values. However, these metrics tell us about capability of the proposed approach to discriminate classes but not about data distribution among classes³³ ³⁴. To robustly characterize a Parkinson digital biomarker is then demanding to explore more robust statistical alternatives that evidence the generalization of the embedded descriptor and estimate the performance for new samples (Figure 3 illustrates typical limitations of standard classification metrics for unseen data being positioned on unknown places). In fact, we hypothesize that Parkinson and control distributions, observed from an embedding representation, should remain with equal properties from training and test samples. To address such assumption, in this work is explored two statistical properties to validate the shape and variance of motor population distributions:

4.4.1. Variance analysis from Homoscedasticity Here, equality among variance of data distributions is estimated through homoscedasticity operators. Particularly, this analysis is carried out for two independent groups $\langle \mathbf{k} \rangle, \langle \mathbf{u} \rangle$ with cardinality $|\mathbf{x}_{\langle \mathbf{k} \rangle}^{(i)}|, |\mathbf{x}_{\langle \mathbf{u} \rangle}^{(j)}|$ of classes $(\mathbf{i}), (\mathbf{j})$. Here, it was considered two dispersion metrics regarding the Levene mean ($\Delta_\ell^{(g)} = |\mathbf{x}_\ell^{(g)} -$

³³ Janez Demšar. “Statistical comparisons of classifiers over multiple data sets”. In: *The Journal of Machine Learning Research* 7 (2006), pp. 1–30.

³⁴ Julián Luengo, Salvador García, and Francisco Herrera. “A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests”. In: *Expert Systems with Applications* 36.4 (2009), pp. 7798–7808.

$\mu(\mathbf{x}^{(g)})$), and the Brown-Forsythe median ($\Delta_\ell^{(g)} = |\mathbf{x}_\ell^{(g)} - med(\mathbf{x}^{(g)})|$). From such dispersion distances, the test statistic W between $\mathbf{x}_{(k)}^{(i)}$ and $\mathbf{x}_{(u)}^{(j)}$ can be defined as:

$$W = \frac{N - |\mathbf{P}|}{|\mathbf{P}| - 1} \frac{\sum_{g \in \mathbf{P}} [|\mathbf{x}_{(g)}| (\mu(\Delta^{(g)}) - \mu(\Delta))^2]}{\sum_{g \in \mathbf{P}} [\sum_{\ell \in \mathbf{x}_{(g)}} (\Delta_\ell^{(g)} - \mu(\Delta^{(g)}))^2]} \quad (4)$$

where $P = \{\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)}, \dots\}$ is the union set of every data group from all classes, $|\mathbf{P}|$ is the cardinality of \mathbf{P} , N is the sum of all $|\mathbf{x}_{(g)}|$ cardinalities, $\mu(\Delta^{(g)})$ correspond to the mean $\langle g \rangle$ of $\Delta_\ell^{(g)}$ values and $\mu(\Delta)$ is the overall mean of every $\Delta_\ell^{(g)}$ value in \mathbf{P} . This estimation evaluate if the samples between two different groups are equally in variance for the same class, leading us to the first step in model generalization for any new sample related to trained data. Additionally the homocedasticity property is useful when is needed to check if two groups remains in the same distribution range, because two distribution can have the same shape (frequency) but be placed at different domain range, indicating a weakness for the model in new data domains.

From a statistical test perspective, the value W rejects the null hypothesis of homocedasticity when $W > f_{\alpha, |\mathbf{P}|-1, N-|\mathbf{P}|}$ where $f_{\alpha, |\mathbf{P}|-1, N-|\mathbf{P}|}$ is the upper critical value of Fischer distribution with $|\mathbf{P}|-1$ and $N-|\mathbf{P}|$ degrees of freedom at a significance level of α (generally 5%). This metric allows to estimate the clustering level for the model and determining if a new data samples from another domains are contained in data distributions of control or Parkinson patients. Then, the homocedasticity value of $\mathbf{x}_{(k)}^{(i)}$ against $\mathbf{x}_{(u)}^{(j)}$ is defined as follow:

$$H(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)}) = \begin{cases} \frac{W(\mu(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})) + W(med(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)}))}{2} & i = j \wedge k \neq u \\ 0 & i = j \wedge k = u \\ \frac{2 - (W(\mu(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})) + W(med(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})))}{2} & i \neq j \end{cases} \quad (5)$$

4.4.2. Shapeness analysis from ChiSquare Here, we quantify the “shapeness” focused in having equally distributions. Following the ChiSquare test χ^2 between $\mathbf{x}_{(k)}^{(i)}$ and $\mathbf{x}_{(u)}^{(j)}$ as:

$$\chi^2 = \sum_{\ell} \frac{(\mathbf{x}_{\ell}^{(k)} - \mathbf{x}_{\ell}^{(u)})^2}{\mathbf{x}_{\ell}^{(u)}} \quad (6)$$

From this rule, it should be considered that both groups must have the same cardinality ($|\mathbf{x}_{(k)}| = |\mathbf{x}_{(u)}|$) and the respective data sorting determines the direction of comparison (*i.e.* the direction goes from group $\langle \mathbf{k} \rangle$ to have the same distribution of $\langle \mathbf{u} \rangle$). To address these issues we make that the lower group will be repeated in its elements without adding new unknow data to preserve its mean and standard deviation, and secondly, we evaluate both directions to quantify the similarity when $\chi^2(\mathbf{x}_{(k)}^{(i)} \rightarrow \mathbf{x}_{(u)}^{(j)})$ and $\chi^2(\mathbf{x}_{(u)}^{(j)} \rightarrow \mathbf{x}_{(k)}^{(i)})$.

The value χ^2 reject the null hypothesis of equal distributions when $\chi^2 > \chi_{\alpha, |\mathbf{x}_{(g)}| - 1}^2$ where $\chi_{\alpha, |\mathbf{x}_{(g)}| - 1}^2$ is the upper critical value of Chi Square distribution with $|\mathbf{x}_{(g)}| - 1$ degrees of freedom at a significance level of α . We define the shapeness value as:

$$Sh(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)}) = \begin{cases} \frac{\chi^2(\mathbf{x}_{(k)}^{(i)} \rightarrow \mathbf{x}_{(u)}^{(j)}) + \chi^2(\mathbf{x}_{(u)}^{(j)} \rightarrow \mathbf{x}_{(k)}^{(i)})}{2} & i = j \wedge k \neq u \\ 0 & i = j \wedge k = u \\ \frac{2 - (\chi^2(\mathbf{x}_{(k)}^{(i)} \rightarrow \mathbf{x}_{(u)}^{(j)}) + \chi^2(\mathbf{x}_{(u)}^{(j)} \rightarrow \mathbf{x}_{(k)}^{(i)})}{2} & i \neq j \end{cases} \quad (7)$$

This test can be used directly as indicator of how relatively far are the samples from each other. Hence, a higher value of this metric means that the samples will be clearly different and separated, but there is the possibility that control patients distribution is near to parkinson's while parkinson can be clearly far. Finally, in algorithm 1 is showed the steps to calculate the proposed homocedasticity and shapeness level for the model.

Algorithm 1 Calculation of homocedastic and shapeness metric for any quantity of data groups with any classes

Require: $C = \{c_0, c_1, \dots, c_n\}$ ▷ Classes in dataset

Require: $G_{c_i} = \{\mathbf{x}_{(0)}^{(i)}, \mathbf{x}_{(1)}^{(i)}, \dots, \mathbf{x}_{(m_i)}^{(i)}\} \forall c_i \in C$ ▷ Partitions per classes

$h \leftarrow 0$

$s \leftarrow 0$

for any pair (c_i, c_j) in C **do**

for any pair $(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})$ in $\cup(G_{c_i}, G_{c_j})$ **do**

$h \leftarrow h + H(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})$ ▷ H defined in eq. 5

$s \leftarrow s + Sh(\mathbf{x}_{(k)}^{(i)}, \mathbf{x}_{(u)}^{(j)})$ ▷ Sh defined in eq. 7

end for

end for

$N \leftarrow \sum_i^n |G_{c_i}|$

$d \leftarrow \binom{N}{2}$ ▷ Combinatory of N in groups of 2

$h \leftarrow \frac{h}{d}$ ▷ Homocedasticity level metric

$s \leftarrow \frac{s}{d}$ ▷ Shapeness level metric

5. EXPERIMENTAL SETUP

5.1. Datasets

5.1.1. Captured dataset In this study were recruited 37 patients from control (23 subjects with average age of 64.7 ± 13) and parkinson (14 subjects with an average age of 72.8 ± 6.8) population. The patients were invited to walk (without any markers protocol), developing a natural locomotion gesture. Parkinson participants were evaluated by a physiotherapist (with more than five years of experience) and stratified according to the H&Y scale (level 1.0 = 2, level 1.5 = 1, level 2.5 = 5, and level 3.0 = 6 participants). These patients written an informed consent and the total dataset count with the approval of the Ethics Committee of Universidad Industrial de Santander.

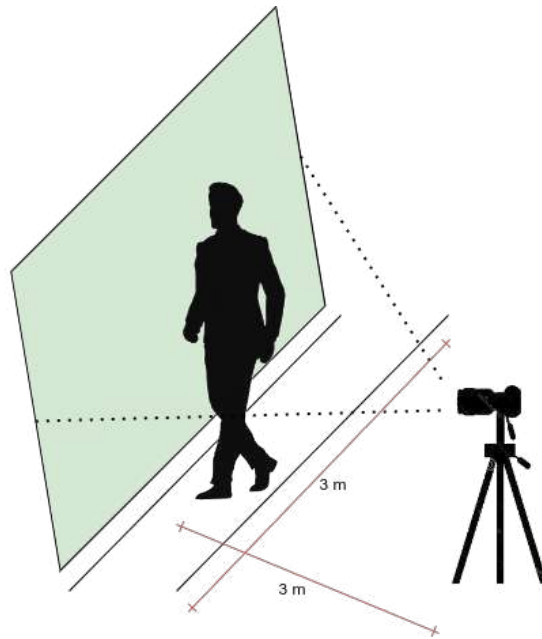


Figure 4. Illustration of the capture setup used for recording each patient developing the gait in front of the camera.

For recording, during a natural walking in around 3 meters, the locomotion was registered 8



Figure 5. Some examples of the recorded patients using the capture setup for our dataset. The first row are control patients and the second parkinson ones.

times from a sagittal view, following a semi-controlled conditions (a green background). In this study we use a conventional optical camera Nikon D3500, that output sequences at 60 fps with a spatial resolution of 1080p. The camera was localized to cover the whole participant silhouette. Every sequence was spatially resized to 64×64 pixels, and temporally cropped to 64 frames. Besides, the videos were normalized and a subsequent subsampling was carried out to ensure a complete gait cycle. To follow one learning class, the proposed approach was trained only with control subjects. In such case, the set of control patients was split in common train, validation and test partitions of 11, 3 and 9 randomly patients selected, respectively. For parkinson participants, we take for validation and test partitions of 3 and 11 patients randomly selected to complement validation and test control sets. Hence, we balanced data for standard and statistical validation purposes.

5.1.2. External dataset validation A main interest in this work is to measure the capability to generalize motion patterns from anomaly deep representations. Also, we are interested in measure the capability of embedding descriptors to discriminate PD from other classes, even

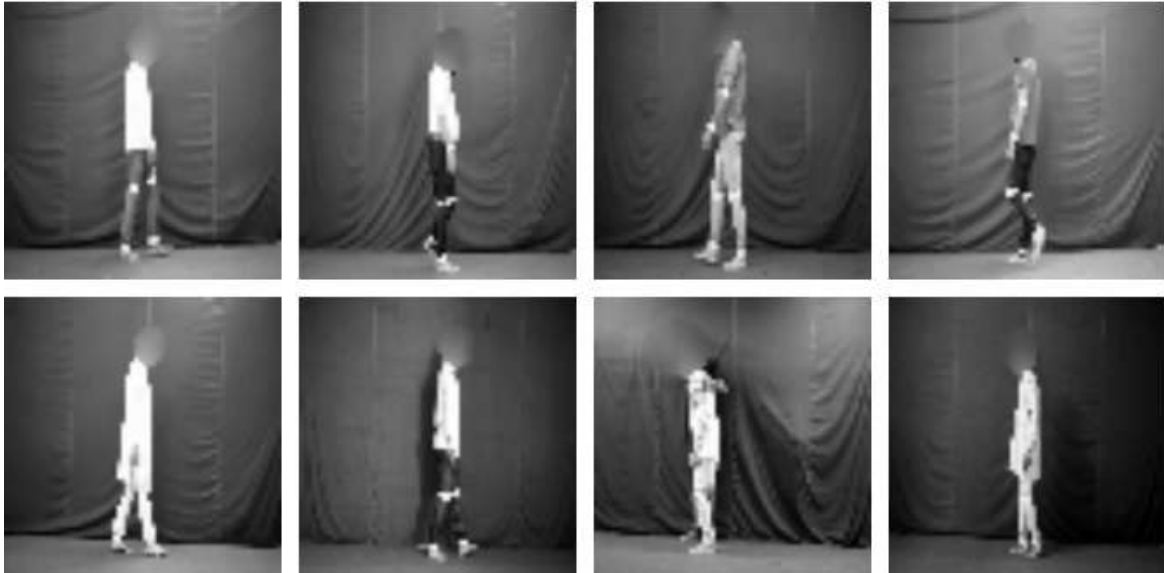


Figure 6. Some examples of the external dataset patients used for validate the model. The first row are control patients and the second parkinson ones. For Knee-Ostheoarthritis patients the videos are similar but not showed.

for videos captured with external protocols. Hence, in this work we only evaluate the proposed approach with a public dataset of walking videos that include knee-osteoarthritis (50 subjects with an average age of 56.7 ± 12.7), parkinson (16 subjects with an average age of 68.6 ± 8.3) and control (30 subjects with an average age of 43.7 ± 9.3) patients³⁵. The 96 participants were recorded with a static green background, blurred faces and markers on their bodies. Following the same methodology for owner data, each sequence was spatially resized to 64×64 pixels, and temporally cropped to 64 frames, and finally normalized and subsampled ensuring a complete gait cycle.

³⁵ Navleen Kour, Sakshi Arora, et al. “A Vision-Based Gait Dataset for Knee Osteoarthritis and Parkinson’s Disease Analysis with Severity Levels”. In: *International Conference on Innovative Computing and Communications*. Springer. 2022, pp. 303–317.

Table 1. Generator and Discriminator Networks structure summary

Module	Network	Levels	Input	Output
Generator	Encoder	5	$64 \times 64 \times 64 \times 1$	$1 \times 1 \times 1 \times n$
	Decoder	5	$1 \times 1 \times 1 \times n$	$64 \times 64 \times 64 \times 1$
Discriminator	Encoder	5	$64 \times 64 \times 64 \times 1$	$1 \times 1 \times 1 \times 1$

5.2. Model configuration

The introduced strategy has in the generator an autoencoder and encoder net, while the discriminator have an encoder net. The encoders use three layers that include 3D ($4 \times 4 \times 4$ and stride $2 \times 2 \times 2$) convolutions, BatchNormalization (momentum of 0.1 and epsilon of 1×10^{-5}) and LeakyRelu ($\alpha = 0.2$). At each progressive level, the input is reduced to half in spatial and temporal dimensions while the features are increased twice. The decoder network follow a symmetrical configuration against the encoder with same layers than encoder (replacing 3D convolutions by 3D transpose convolutions). The overall structure is summarized in table 1.

6. EVALUATION AND RESULTS

6.1. Captured dataset results

The proposed strategy was exhaustively validated with respect to the capability to recognize parkinsonian inputs as abnormal class patterns in architectures trained only with control patterns and under challenging unbalanced and scarce scenarios. Hence, in the first experiment, the proposed strategy was trained only with control samples from owner dataset, following a video reconstruction pretext task. Hence, encoder ($\|\mathbf{z}_G - \mathbf{z}'_G\|_2$), contextual ($\|\mathbf{x} - \mathbf{x}'\|_1$) and adversarial ($\|\mathbf{z}_D - \mathbf{z}'_D\|_2$) embedding errors were recovered as locomotor descriptors of the observed sequences. For classification purposes, these errors were binarized by imposing a threshold value, as: $\tau_{\mathbf{z}_G} = 1.768$ for encoder, $\tau_{\mathbf{x}} = 0.147$ for contextual, and $\tau_{\mathbf{z}_D} = 0.429$ for adversarial errors. Table 2 summarizes the achieved performance of three locomotor descriptors according to standard classification metrics. In general, the proposed strategy reports a remarkable capability to label parkinson patterns as abnormal samples, which are excluded from trained representation. Interestingly, the contextual errors have the highest value among the others to classify between control and parkinson patients, reporting a remarkable 86.9% in AUC, with mistakes in only 64 video clips (approximately 3 patients).

Table 2. Model performance for encoder, contextual and adversarial losses using standard metrics when the model trains with control patients. Acc, Pre, Rec, Spe, F1 are for accuracy, precision, recall, specificity and f1 score respectively.

Loss	Acc	Pre	Rec	Spe	F1	ROC-AUC
Encoder	53.8%	89.5%	20.4%	96.9%	33.2%	58.7%
Contextual	85.7%	96.6%	77.4%	96.4%	85.7%	86.9%
Adversarial	75.5%	94.3%	60%	95.4%	73.3%	77.7%

6.2. Statistical validation of captured dataset

For robustness validation, we are also interested in the distribution output of predictions, which may suggest the capability of generalization of the model. For doing so, we also validate locomotion descriptors with respect to introduced homoscedasticity and shapeness validation. Table 3 summarizes the results achieved by each locomotion embedding descriptor, contrasting with the reported results from standard metrics. In such case, the validated metrics suggest that contextual errors may be overfitted for the trained dataset and the recording conditions, which may be restrictive for generalized architecture in other datasets. Contrary, the encoder descriptor shows evident statistical robustness from variance and shapeness distributions. Furthermore, the encoder losses evidence a clearly separation between the control and parkinson distribution in Figure 7, where even the proposed model can separate stages of Hoehn & Yahr with the difference between 2.5 and 3.0 levels where the ChiSquare test show us that both distribution remains equals meaning that both stages are difficult to model.

Table 3. Model performance for encoder, contextual and adversarial losses using the proposed statistical metrics when the model trains with control patients.

Loss	Homocedasticity	Shapeness
Encoder	80%	70%
Contextual	50%	40%
Adversarial	50%	45%

6.3. External dataset results

To follow with one of the main interests in this work *i.e.*, the generalization capability, the proposed strategy was validated with an external public dataset (without any extra training) that include parkinson (16 patients), knee-osteoarthritis (50 patients) and control patients (30

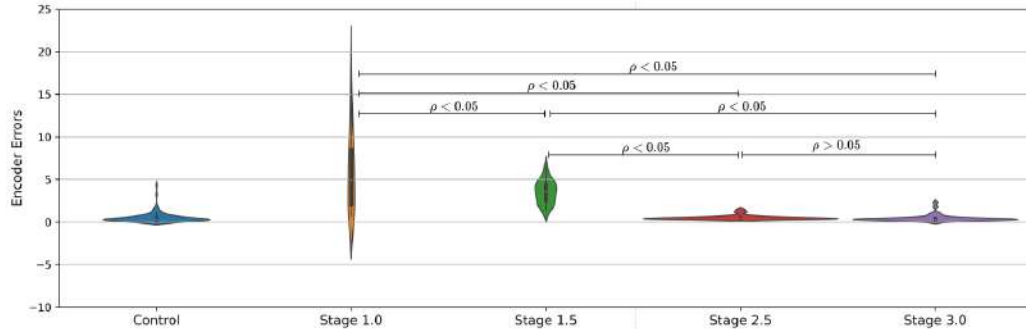


Figure 7. Data distribution given by the proposed model for control and parkinson samples by Hoehn & Yahr levels.

patients) ³⁵. Table 4 summarized the achieved results to discriminate among the three unseen classes, evidencing a notable performance following encoder embedding representation. It should be noted, that Encoder achieves the highest ROC-AUC, reporting an average of 75%, being the more robust representation, as suggested by statistical homoscedasticity and shapeness validation. The contextual and the adversarial losses have better accuracy, precision and recall, but the specificity suggest that there is not any evidence of correctly classifying control subjects. In such sense, the model label all samples as abnormal from trained representation. In contrast, the encoder element in the network (Figure 2-a) capture relevant gait patterns to distinguish between control, parkinson and knee-osteoarthritis patients.

Table 4. Model performance for encoder, contextual and adversarial losses using the proposed model without retraining and same thresholds as Table 2. Acc, Pre, Rec, Spe, F1 are for accuracy, precision, recall, specificity and f1 score respectively.

Loss	Acc	Pre	Rec	Spe	F1	ROC-AUC
Encoder	62.6%	97.9%	58.1%	91.9%	72.9%	75%
Contextual	86.7%	86.7%	100%	0%	92.9%	50%
Adversarial	87.8%	89.4%	97.4%	24.9%	93.3%	61.2%

6.4. Statistical validation of external dataset

Along the same line, the external dataset was also validated with respect to homoscedasticity and shapeness metrics. Table 5 summarizes the achieved results from the distribution representation of output probabilities. As expected, the results enforce the fact that embeddings from the Encoder have much better generalization against the other losses, allowing to discriminate among three different unseen classes. Remarkably, the results suggest that control subjects of the external dataset belong to the trained control set. This fact is relevant because indicates that architecture is principally dedicated to coded locomotor patterns without strict restrictions about captured conditions. To complement such results, output probabilities from three classes are summarized in violin plots, as illustrated in Figure 8 where shows the separation between the classes of parkinson and knee-osteoarthritis, also, between levels of the diseases, being remarkable the locomotor affectations produced by the patients diagnosed with knee-Osteoarthritis.

Table 5. Model performance for encoder, contextual and adversarial losses using the proposed statistical metrics and model as Table 2.

Loss	Homocedasticity	Shapeness
Encoder	66.7%	66.7%
Contextual	83.4%	0%
Adversarial	16.7%	16.7%

6.5. Parkinson training validation

Alternatively, in an additional experiment we train using only patients diagnosed with parkinson to force the architecture to extract these abnormal locomotion patterns. In such cases, the videos from control subjects are associated with abnormal responses from trained archi-

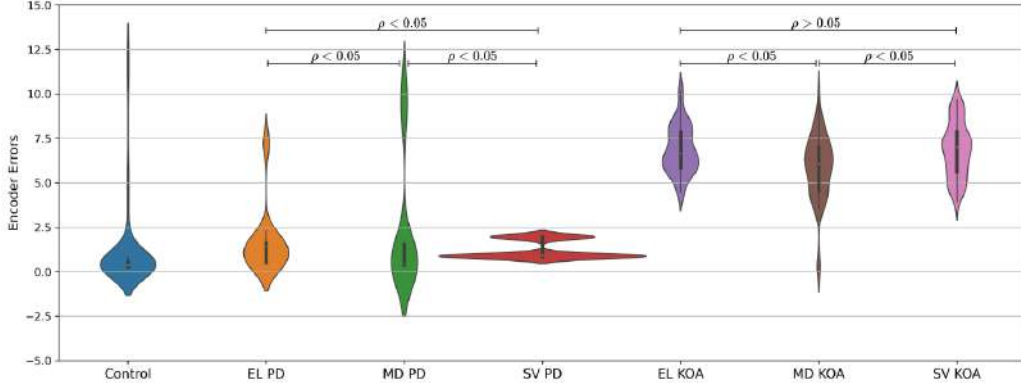


Figure 8. Data distribution given by the proposed model for control, parkinson (PD) and knee-osteoarthritis (KOA) samples by levels where EL is early, MD medium and SV severe.

Table 6. Model performance for encoder, contextual and adversarial losses using standard metrics when the model trains with parkinson patients. Acc, Pre, Rec, Spe, Homo and Shape are for accuracy, precision, recall, specificity, homocedasticity and shapeness respectively.

Loss	Acc	Pre	Rec	Spe	Homo	Shape	ROC-AUC
Encoder	62.5%	55.2%	88.9%	40.9%	45%	50%	64.9%
Contextual	71.5%	93.5%	73.7%	50%	50%	40%	61.9%
Adversarial	68.8%	64.1%	69.4%	68.2%	45%	40%	68.8%

ture. Table 6 summarizes the achieved results from standard and statistical distribution metrics. As expected, from this configuration of the architecture is achieved a lower classification performance because the high variability and complexity to code the disease. In fact, parkinson patients may manifest totally different locomotion affectations at the same stage. For such reason, the architecture has major challenges to discriminate control subjects and therefore lower agreement with ground truth labels. The statistical homocedasticity and shapeness metrics confirm such issue achieving scores lower than 50% and indicating that the model, from such configuration, is not generalizable. In this configuration, it would be demanding a larger amount of parkinson patients to deal with disease variability.

7. DISCUSSION

This work presented a deep generative scheme, designed under the one-class-learning methodology to model gait locomotion patterns in markerless video sequences. The proposed architecture is trained under the reconstruction video pretext task, being categorical to capture kinematic behaviors without the association of expert diagnosis criteria. From an exhaustive experimental setup, the proposed approach was trained with videos recorded from a control population, while then parkinsonian patterns were associated with anomaly patterns from the design of a discrimination metric that operates from embedding representations. From an owner dataset, the proposed approach achieves an ROC-AUC of 86.9%, while for an external dataset without unseen training videos, the proposed approach achieved an average ROC-AUC of 75%.

One of the main issues addressed in this work was to make efforts to train generative architecture with a sufficient generalization capability to capture kinematic patterns without a bias associated to the capture setups. To carefully select such architectures, this study introduced homoscedasticity and shapeness as complementary statistical rules to validate the models. From these metrics was evidenced that encoder embeddings brings major capabilities to generalize models, against the contextual and adversarial losses, achieving in average an 80% and 70% for homocedasticity and shapeness, respectively. Once these metrics defined the best architecture and embedding representation, we confirm the selection by using the external dataset with different capture conditions and even with the study of a new disease class into the population *i.e.*, the Knee-osteoarthritis. Remarkably, the proposed approach generates embeddings with sufficient capabilities to discriminate among different unseen populations.

In the literature have been declared different efforts to develop computational strategies to discriminate parkinson from control patterns, following markerless and sensor-based observations

^{6 7 8 9 36}. For instance, volumetric architectures have been adjusted from discriminatory rules taking minimization rules associated with expert diagnosis annotations ^{6 8}. These approaches have reported remarkable results (average an 95% ROC-AUC with 22 patients). Also, Sun *et. al.* proposed an architecture that takes frontal gait views and together with volumetric convolution layers, discriminates the level of freeze in the gait for parkinson patients with an accuracy of 79.3%. Likewise, Kour *et. al.* ³⁶ develops a sensor-based approach to correlate postural relationships with several annotated disease groups (reports an accuracy = 92.4%, precision = 90.0% with 50 knee-osteoarthritis, 16 parkinson and 30 control patients). Nonetheless, such schemes are restricted to a specific recording scenario and pose observational configurations. Besides, the minimization of these representations may be biased by label annotations associated with expert diagnostics. Contrary, the proposed approach adjusts the representation using only control video sequences without any expert label intervention during the architecture tuning. In such case, the architecture has major flexibility to code potential hidden relationships associated with locomotor patterns. In fact, the proposed approach was validated with raw video sequences, reported in ³⁶, surpassing precision scores without any additional training to observe such videos. Moreover, the proposed approach uses video sequences instead of representation from key points, that coarsely minimize dynamic complexity during locomotion.

Recovered generalization metrics scores (homocedasticity = 80%, shapeness = 70%) suggest that some patients have different statistical distributions, an expected result from variability in control population, as well as, the variability associated to disease parkinson phenotyping. In such sense, it is demanding a large set of training data to capture additional locomotion components, together with a sufficient variability spectrum. Nonetheless, the re-training of the architecture should be supervised from output population distributions to avoid overfitting regarding specific training scenarios. The output reconstruction may also be extended as anomaly

³⁶ Navleen Kour, Sunanda Gupta, and Sakshi Arora. “A vision-based clinical analysis for classification of knee osteoarthritis, Parkinson’s disease and normal gait with severity based on k-nearest neighbour”. In: *Expert Systems* 39.6 (2022), e12955.

maps to evidence in the spatial domain the regions with anomalies, which further may represent some association with the disease to help experts in the correct identification of patient prediction.

8. CONCLUSIONS AND FUTURE WORK

This work presented a deep generative architecture with the capability of discovering anomaly locomotion patterns, convolving entire video sequences into a 3D scheme. Interestingly, a parkinson disease population was projected to the architecture, returning not only outlier rejection but coding a new locomotion distribution with separable patterns with respect to the trained control population. These results evidenced a potential use of this learning and architecture scheme to recover potential digital biomarkers, coded into embedding representations. The proposed approach was validated with classical classification rules but also with statistical measures to validate the capability of generalization.

Future works include the validation of proposals among different stages and the use of federated scenarios with different experimental capture setups to test performance on real scenarios. The proposed approach evidence remarkable results following a reconstruction pretext task. As a perspective, complementary auxiliary tasks may be evaluated to capture hidden locomotion patterns without explicit annotations. Also, in perspective of this line of work is expected to explore the capability of statistical metrics to recognize generality capabilities in more general scenarios and in other extended applications.

BIBLIOGRAPHY

- Akçay, Samet, Amir Atapour-Abarghouei, and Toby P Breckon. “Ganomaly: Semi-supervised anomaly detection via adversarial training”. In: *Asian conference on computer vision*. Springer. 2018, pp. 622–637 (cit. on p. 20).
- Akçay, Samet, Amir Atapour-Abarghouei, and Toby P Breckon. “Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection”. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019, pp. 1–8 (cit. on p. 20).
- Balestrino, R and AHV Schapira. “Parkinson disease”. In: *European journal of neurology* 27.1 (2020), pp. 27–42 (cit. on p. 11).
- Chalapathy, Raghavendra and Sanjay Chawla. “Deep learning for anomaly detection: A survey”. In: *arXiv preprint arXiv:1901.03407* (2019) (cit. on pp. 16–19).
- Chandola, Varun, Arindam Banerjee, and Vipin Kumar. “Anomaly detection: A survey”. In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on p. 16).
- Chen, Shih-Wei et al. “Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis”. In: *Biomedical engineering online* 10.1 (2011), pp. 1–21 (cit. on p. 18).
- Cho, Chien-Wen et al. “A vision-based analysis system for gait recognition in patients with Parkinson’s disease”. In: *Expert Systems with applications* 36.3 (2009), pp. 7033–7039 (cit. on p. 18).
- Demšar, Janez. “Statistical comparisons of classifiers over multiple data sets”. In: *The Journal of Machine Learning Research* 7 (2006), pp. 1–30 (cit. on p. 28).

- Di Mattia, Federico et al. “A survey on gans for anomaly detection”. In: *arXiv preprint arXiv:1906.11632* (2019) (cit. on p. 19).
- Dorsey, E Ray and Bastiaan R Bloem. “The Parkinson pandemic—a call to action”. In: *JAMA neurology* 75.1 (2018), pp. 9–10 (cit. on p. 11).
- Goetz, Christopher G et al. “Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results”. In: *Movement disorders: official journal of the Movement Disorder Society* 23.15 (2008), pp. 2129–2170 (cit. on p. 12).
- Goodfellow, Ian J et al. “Generative adversarial networks”. In: *arXiv preprint arXiv:1406.2661* (2014) (cit. on p. 16).
- Guayacán, Luis C and Fabio Martínez. “Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network”. In: *Journal of biomedical informatics* 123 (2021), p. 103935 (cit. on pp. 12, 18, 42).
- Guayacán, Luis Carlos, Edgar Rangel, and Fabio Martínez. “Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 3688–3691 (cit. on pp. 12, 18, 42).
- Han, Xu, Xiaohui Chen, and Li-Ping Liu. “GAN Ensemble for Anomaly Detection”. In: *arXiv preprint arXiv:2012.07988* (2020) (cit. on p. 20).
- Kiran, B Ravi, Dilip Mathew Thomas, and Ranjith Parakkal. “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos”. In: *Journal of Imaging* 4.2 (2018), p. 36 (cit. on p. 18).

- Kour, Navleen, Sakshi Arora, et al. “A Vision-Based Gait Dataset for Knee Osteoarthritis and Parkinson’s Disease Analysis with Severity Levels”. In: *International Conference on Innovative Computing and Communications*. Springer. 2022, pp. 303–317 (cit. on pp. 34, 38).
- “Computer-vision based diagnosis of Parkinson’s disease via gait: a survey”. In: *IEEE Access* 7 (2019), pp. 156620–156645 (cit. on pp. 12, 17, 18).
- Kour, Navleen, Sunanda Gupta, and Sakshi Arora. “A vision-based clinical analysis for classification of knee osteoarthritis, Parkinson’s disease and normal gait with severity based on k-nearest neighbour”. In: *Expert Systems* 39.6 (2022), e12955 (cit. on p. 42).
- Li, Michael H et al. “Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation”. In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), pp. 1–13 (cit. on pp. 12, 18, 42).
- Litjens, Geert et al. “A survey on deep learning in medical image analysis”. In: *Medical image analysis* 42 (2017), pp. 60–88 (cit. on pp. 12, 14–16).
- Luengo, Julián, Salvador García, and Francisco Herrera. “A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests”. In: *Expert Systems with Applications* 36.4 (2009), pp. 7798–7808 (cit. on p. 28).
- Nayak, Rashmiranjan, Umesh Chandra Pati, and Santos Kumar Das. “A comprehensive review on deep learning-based methods for video anomaly detection”. In: *Image and Vision Computing* 106 (2021), p. 104078 (cit. on pp. 19, 20).
- Nõmm, Sven et al. “An alternative approach to distinguish movements of parkinson disease patients”. In: *IFAC-PapersOnLine* 49.19 (2016), pp. 272–276 (cit. on p. 18).

- Organization, World Health. *Neurological disorders: public health challenges*. World Health Organization, 2006 (cit. on p. 11).
- Ravanbakhsh, Mahdyar et al. “Abnormal event detection in videos using generative adversarial nets”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 1577–1581 (cit. on p. 19).
- Ravanbakhsh, Mahdyar et al. “Training adversarial discriminators for cross-channel abnormal event detection in crowds”. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2019, pp. 1896–1904 (cit. on p. 20).
- Schlegl, Thomas et al. “f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks”. In: *Medical image analysis* 54 (2019), pp. 30–44 (cit. on p. 20).
- Schlegl, Thomas et al. “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery”. In: *International conference on information processing in medical imaging*. Springer. 2017, pp. 146–157 (cit. on p. 19).
- Schmarje, Lars et al. “A survey on semi-, self-and unsupervised learning for image classification”. In: *IEEE Access* 9 (2021), pp. 82146–82168 (cit. on p. 18).
- Soltaninejad, Sara et al. “Body movement monitoring for Parkinson’s disease patients using a smart sensor based non-invasive technique”. In: *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE. 2018, pp. 1–6 (cit. on p. 18).
- SPLETE, HEIDI. “Avoid Diagnostic Pitfalls For Parkinson’s Disease”. In: *Internal Medicine News* 39.17 (2006), pp. 31–31 (cit. on p. 21).

- Sugandhi, K, Farha Fatina Wahid, and G Raju. “Feature extraction methods for human gait recognition—a survey”. In: *International Conference on Advances in Computing and Data Sciences*. Springer. 2016, pp. 377–385 (cit. on pp. 18, 21).
- Sun, Renfei et al. “Convolutional 3D attention network for video based freezing of gait recognition”. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE. 2018, pp. 1–7 (cit. on pp. 12, 18, 42).
- Vos, Theo et al. “Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”. In: *The Lancet* 390.10100 (2017), pp. 1211–1259 (cit. on p. 11).
- Zenati, Houssam et al. “Efficient gan-based anomaly detection”. In: *arXiv preprint arXiv:1802.06222* (2018) (cit. on p. 20).

APPENDICES

Anexo A. Academic Products

Posters

- Rangel, E., & Martínez, F. (2022, November). "A Parkinsonian Digital Biomarker Learned as an Anomaly Deep Generative Representation". In 2022 International Workshop on Adaptive, Compressive and Computational Imaging (WACCI 2022). Universidad Industrial de Santander
Status: Presented.

Journals

- Rangel, E., & Martínez, F. (2022) "Parkinson gait modelling from an anomaly deep representation". Pattern Recognition, 2022. United Kingdom.
Status: Submitted.

Conference papers

- Rangel, E., & Martínez, F. (2022, July). "A Parkinsonian Digital Biomarker Learned as an Anomaly Deep Generative Representation". In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 4188-4191). IEEE.
Status: Presented.

Collaborations

- Guayacán, L. C., Rangel, E., & Martínez, F. (2020, July). "Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network". In

2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3688-3691). IEEE.

Status: Published.

- Rodríguez, J., Chacón, J., Rangel, E., Guayacán, L., Hernández, C., Hernández, L., & Martínez, F. (2020). "Understanding Motion in Sign Language: A New Structured Translation Dataset". In Proceedings of the Asian Conference on Computer Vision.

Status: Published.

- Gómez, S., Mantilla, D., Rangel, E., Ortiz, A., DVera, D., & Martínez, F. (2022). "A deep supervised cross-attention strategy for ischemic stroke segmentation in MRI studies". IEEE Transactions on Biomedical Engineering, 2022. United States.

Status: Submitted.