FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Miss SAIGON – Missing Signal Appraising in Globally Optimized Networks

Luís Miguel Brito Teixeira

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Professor Doutor Vladimiro Henrique Barrosa Pinto de Miranda

Second Supervisor: Professor Doutor Jorge Pereira

July 26, 2019

Resumo

O tipo de intervenientes na rede de energia mudou desde o estabelecimento das linhas de transmissão e distribuição até os dias de hoje. Com a integração das fontes renováveis e as atuais condições variáveis do mercado, as condições de operação do sistema são mais restritivas, de forma a garantir o fornecimento contínuo de energia. No entanto, o operador do sistema não pode observar todos os eventos na rede devido à falta de observabilidade do sistema ou o evento ocorre sem alerta-lo.

Além da existência de mais PMUs do que antes, o número delas não é relevante, e a possibilidade de falhas na comunicação de dados também é uma preocupação. A fim de fornecer um reconhecimento adequado da topologia da rede, a presente dissertação define um processador de topologia único baseado numa estrutura de *Deep Learning*, a *Convolutional Neural Network* (CNN). Além disso, os conceitos da teoria da informação são usados para medir o quanto uma variável de topologia está correlacionada à conectividade do disjuntor longínquo. Ambas as duas áreas são importantes para definir um processador de topologia correto para fornecer as informações de topologia para o operador do sistema com eficiência.

Um cenário de operação realista de poucas medidas disponíveis é apresentado com uma classificação impressionante do estado do interruptor. Além disso, o problema de determinação da topologia de subestação é tratado aqui com uma nova abordagem do problema. Esta dissertação além de contribuir para uma correta determinação da topologia da rede, proporcionando também um planeamento da instalação ótima de medidores e da PMU.

Abstract

The type of network intervenients changed since the establishment of the transmission and distribution lines until nowadays. With the integration of renewable sources and the actual variable market conditions, the conditions of the system operation are more restrictive, in order to guarantee continuous energy supply. However, the system operator cannot observe all the events on the grid due to the lack of system observability, or the event occurs without alerting the system operator.

Besides the existence of more PMUs than before, the number of them is not relevant, and the possibility of failed data communication is also a concern. In order to provide a proper acknowledgement of the network topology, the present dissertation defines a unique topology processor based on a Deep Learning framework, the Convolutional Neural Network (CNN). Also, information theory concepts are used in order to measure how much a topology variable is correlated to the remote breaker connectivity. Both two areas are important to define a correct topology processor to provide the topology information to the system operator efficiently.

A realistic operation scenario of a few available measurements is presented with an impressive breaker status classification. Also, the substation topology determination problem is addressed here with a new concern approach. This dissertation beyond contributes to a correct determination of network topology, also providing a planning of meters and PMU optimal installation.

Keywords: Information Theory Learning, Convolutional Neural Networks, Deep Learning, topology processor, breaker status, substation topology, meter, PMU.

Acknowledgements

First of all, I would like to thank my dissertation supervisor, Prof. Dr Vladimiro Miranda for all the support through this adventure. This work could not be done without his clear vision, motivation and inspiring ideas, making me very proud of what was achieved. Dr Jorge Pereira, I also would like to thank you for support.

I would like to address a special thank you to Pedro Cardoso, Francisco Barbosa and Miguel Barros for all discussed and inspiring ideas.

For Inês, I would like to express all my gratitude for the constant support on this stage of my life and all the shared love and motivation. In addition, I am grateful for your contribution to towards improving my english.

Finally, to my parents and sister, I address all my love for the values and the opportunity to follow my dreams. Also, to my uncles and friends, Augusto and Ivo, I would like to dedicate this work for all the good memories that we lived and make me feel *saudade*.

Luís Miguel Brito Teixeira

*"Failure is simply the opportunity to begin again,
this time more intelligently."*

Henry Ford

Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives 3
1.3 Dissertation Organisation 4

2 State of the art **5**
2.1 State Estimation 5
2.1.1 Problem overview 5
2.1.2 Minimise Errors - WLS 7
2.1.3 Topology Processing Problem - Classical problem overview 9
2.1.4 Topology Processing Problem - new paradigm 12
2.2 Deep Learning - an uncommon approach 13
2.2.1 Deep Learning vs Classical methodologies 13
2.2.2 CNNs - Convolutional Neuronal Networks 15
2.3 Final Remarks 19

3 ITL - Information Theory Learning **21**
3.1 Brief introduction to Information Theory Concept 21
3.1.1 Shannon's Entropy 21
3.1.2 Renyi's Entropy 22
3.1.3 Renyi's Quadratic Entropy - a particular case 23
3.1.4 Parzen Windows method 24
3.1.5 Distance of Cauchy-Schwarz 25
3.2 How much a measurement defines a state of a breaker 27
3.3 Final Remarks 29

4 Topology processor based on CNN **31**
4.1 Test system - dataset generation 31
4.2 CNN model 33
4.2.1 Classification phase 33
4.2.2 CNN structures 34
4.2.3 Training Procedure 36
4.3 Input structure organisation 40
4.4 Substation internal topology 41
4.5 Final Remarks 43

5 Results **45**
5.1 Initial considerations - Dataset 45
5.2 Single Breaker Analysis 46
5.2.1 Corroboration of models 46
5.2.2 Influence of input arraignment 50
5.2.3 Performance under lack of measurements 54
5.2.4 Realistic point of view 57
5.3 Substation Topology 61
5.4 PMU introduction 64
5.5 Final Remarks 67

6 Conclusions and Future Work **69**
6.1 Conclusions 69
6.2 Future Work 70

A Test System **73**

References **75**

List of Figures

1.1 Example of self-healing structure in order to operate a distribution or transmission network [1]. 2

2.1 Flowcharts representing differences between traditional programming approaches, classical machine learning and, what can it is possible to achieve in the AI field based on machine learning techniques [39]. 14

2.2 Evolution of performance on different common applied types of machine learning with the amount of available data [40]. 14

2.3 Approximation of neuron comportment to apply in neural networks with weights w_i, biases b_i and input x_i [43]. 16

2.4 The most used activation functions in neural networks field: Sigmoid (ON TOP), Hyperbolic Tangent (MIDDLE) and ReLU (DOWN) where the input z is affected by them. 16

2.5 Visual example of a convolution operation with a clear demonstration of sparse interactions, parameter sharing and equivariant representation proprieties with a matrix of weights (kernel), and input image (pixels matrix value representation) [39]. 17

2.6 Demonstration of an example of CNN architecture, LeNet-5 used to digit classification where it is observed the three principal layers: convolutional, polling and fully-connected layer [42]. 18

3.1 Parzen Windows method applied to X with $\sigma = 0.2$. The black dashed Gaussian curves represent pdf of each x_i element and the red curve the pdf estimation following Parzen Windows technique. 25

3.2 Parzen Windows method applied to X with $\sigma = 0.8$. The black dashed Gaussian curves represent pdf of each x_i element and the red curve the pdf estimation following Parzen Windows technique. 25

3.3 Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with p(x) (tiny blue dash), z(x) (red line) and q(x) (strong blue dash) [60]. . 26

3.4 Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with P(X) (red line), $P(X|Y = OFF) \times P(Y = OFF)$ (dashed black line) and $P(X|Y = ON) \times P(Y = ON)$ (dashed light blue line) 28

3.5 Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with P(X) (red line), $P(X|Y = OFF) \times P(Y = OFF)$ (dashed black line) and $P(X|Y = ON) \times P(Y = ON)$ (dashed light blue line) 28

4.1 Representation of load level as pdf in power flow estimation. 32

4.2 Breakers arrangement in the test IEEE RTS 24-bus system. 32

4.3 Proposal Classifier with demonstrative input values to achieve a binary classification. 34

4.4 Layout of CNN structure to 3 layer example with principal operations used in the classification problem of breaker status recognition, ON or OFF. 36
4.5 Illustration of the gradient descendent technique used to a single input function $f(x)$ with path demonstration attaining global minimum on an iterative procedure [39]. 38
4.6 Diagram representing i epochs of iterative procedure with batch size n. 39
4.7 Iterative procedure representing overfitting event along with epochs number increases [39]. 39
4.8 Input structure with 11x11 dimension and organised by D_{CS} criterion where 1^{st} represents measurement with greater distance (most content representation) until to 121^{th} position, value with the lowest distance and less contribution to breaker status definition. 40
4.9 Input structure with 6x6 dimension and organised by D_{CS} criterion where 1^{st} represents measurement with greater distance (most content representation) until to 16^{th} position, value with the lowest distance and less contribution to breaker status definition comparing to 1^{st} value. 41
4.10 Scheme of internal topology breakers of the substation located on bus 9 (LEFT) and bus 15 (RIGHT) with connections to respective buses. 42

5.1 Error accuracy of training procedure for breaker 9 classification using model A and considering 121 available measurements as the non-organise input values. . . 47
5.2 Error accuracy of training procedure for breaker 9 classification using model A and consider 121 available measurements as the non-organise input values. . . . 49
5.3 Error accuracy of training procedure for breaker 9 classification using model A and consider 121 available measurements as the non-organise input values. The input normalisation was made on a range of $[-1,1]$. 50
5.4 Breaker 8 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation. 51
5.5 Breaker 9 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation. 52
5.6 The visual representation of the 16 most valuable measurements to define breaker 9 connectivity. Values were surrounding 0 p.u represents red tonality and module values bigger than it, is linked to yellow tonality. 53
5.7 Representation of values that power flow on line 17-18 can exhibit and comparison with power flow of false close identification scenarios. 53
5.8 Illustrative scheme of input matrix organisation by D_{CS} distance criterion that feds CNN model B and determines breaker 8 status where coloured numbers represent available values and grey tonality the unavailable measurements. 55
5.9 Breaker 1 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation. 56
5.10 Breaker 7 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation. 56
5.11 Breakers arrangement in test IEEE RTS 24-bus system where red lines represent the location of meters. The rest of lines are considered unavailable to performed tests. 58

5.12 Breaker 2 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation. 59
5.13 Breaker 6 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation. 59
5.14 Breaker 1 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation. 60
5.15 Scheme of internal topology breakers of the substation located on bus 15 with connections to respective buses. 62
5.16 Breaker 1 from substation 15 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from substation localisation. 63
5.17 Breaker 3 from substation 15 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from substation localisation. 63
5.18 Scheme of internal topology breakers of the substation located on bus 9 with connections to respective buses. 64

List of Tables

4.1 Specification of developed models, each layer and variable parameters. 35

5.1 Architecture of neural networks model with a different number of free parameters integrated on it. 48
5.2 Comparative results between models with a different number of free parameters integration. 48
5.3 Performance of model A on breaker 9 classification problem with three different neural activation functions: ReLU, hyperbolic tangent and sigmoid. 48
5.4 Reconstruction of the 10 breaker status with 121 values input using Model A and also with an equal non-defined organisation measurement. 50
5.5 Reconstruction of 10 breaker status with 121 values input using Model A and the matrix values organisation mentioned in section 4.3 of chapter 4. 51
5.6 The accuracy results of executed tests under reduction to 16 possibles input values and without direct measurements. 55
5.7 Accuracy results of executed tests under reduction to 16 most informative input values and without direct measurements. 57
5.8 The accuracy results of executed tests under available 8 meters scenario to 10 single breakers. 60
5.9 Application of Model B to breakers from substation 15 and accuracy of each one as the global efficiency of procedure. 62
5.10 Application of Model B to breakers from substation 9 and the accuracy of each one as the global efficiency of the procedure. 64
5.11 CNN model B classification to incorporated switchers on substation 15 with introduction of bus 15 voltage measurements. 65
5.12 CNN model B classification to incorporated switchers on substation 15 with introduction of bus 24 voltage measurements. 65
5.13 CNN model B classification to incorporated switchers on substation 9 with introduction of bus 9 voltage measurements. 65
5.14 CNN model B classification to incorporated switchers on substation 9 with introduction of bus 25 (secondary of substation 9) voltage measurements. 66
5.15 CNN model B classification to incorporated switchers on substation 9 with introduction of best results of bus 25 (secondary of substation 9) voltage measurements experience. 66

A.1 Parameters of IEEE 24-bus test system. 73

List of acronyms and symbols

ADA	Advanced Distribution Automation
AE	Autoencoders
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
DCNN	Deep Convolutional Neural Networks
DMS	Distribution Management System
DSO	Distribution System Operator
EMS	Energy Management System
FSE	Fuzzy State Estimation
GPU	Graphics Processing Unit
GSE	Generalised State Estimation
ITL	Information Theory Learning
LNRT	Largest Normalized Residual Test
MMI	Maximum Mutual Information
MLP	Multilayer Perceptron
NLL	Negative Log-Likelihood
pdf	probability density function
PMU	Phasor Measurement Unit
R-CNN	Region-based Convolutional Neural Networks
ReLU	Rectified Linear Units
TSO	Transmission System Operator
SCADA	Supervisory Control And Data Acquisition
WLS	Weighted Least Squares

Chapter 1

Introduction

The present chapter has the main goal of demonstrating a brief overview of the developed work on this dissertation, guiding the reader to understand the main concern, objectives, solutions to the problem stated and finally the document's organisation.

1.1 Motivation

One of the immediate concerns related to network system operation resides on the evolution that principal stakeholders impose on daily behaviour. The trends of energy consumption are different from the time that distribution lines were projected, and the power supply tends to be done by renewable sources. Nowadays, mainly in Portugal, there is a realistic goal of 100% renewable energy production.

Nowadays, a behaviour change of the network intervenients on old established networks contributes to fast variations of the conditions of operations imposed, for example, by the uncertainty of renewable energy injection on the grid. Also, the change of the actual load supply diagram, requires a higher variation of generation on shorter times, mainly on peak hours. Consequently, from these two main problems, the market condition variability is an actual paradigmatic issue. On the other hand, in order to preserve the main objective of all system operators, the security and reliability of energy supplied should be maintained, and faster control actions have to be incorporated on TSO and DSO. In the last decades, it was observed the automation of network procedures that define a group of Advanced Distribution Automation – ADA – implementation in order to improve some of the next topics:

- The intention to reduce the number of power outages and in the worst case to decrease the recovery time;
- Follow up the integration of distributed generation, calling all the stakeholders to be part of this concept, ranging from the particular utilities to the primary producers;
- Improving the reliability of systems and power distribution quality.

Considering the behaviour trends previously stated, the integration of the renewable source requires the existence of more power electronic devices to complement the traditional generators. These generators are responsible for the system stability due to the faster events incident. A review of the current protections configuration is imperatory, as well as the implemented control actions system. Hence, with the initial implementation of ADA techniques, a unique idea emerges in the beginning of the XXI century, namely, the possible implementation of a **self-healing network**.

The idea of a network without intervention of the system operator is ideal, since it avoids human errors and, most importantly, saves telemetry and human application, executing faster control actions. Nowadays, this concept makes sense on the mentioned operation conditions, and self-healing networks can face most significant problems.

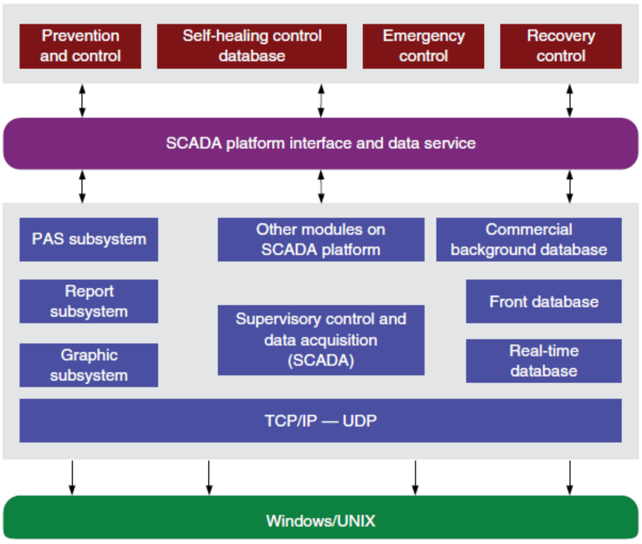

Figure 1.1: Example of self-healing structure in order to operate a distribution or transmission network [1].

Figure 1.1 presents the main clusters that could define a self-healing structure, focusing on the network data acquisition task and on the four main control actions clusters: Prevention and control; Self-healing control database; Emergency control and Recovery control. All of them are dependent of one idea, namely, the knowledge of the network status in order to take adequate control actions.

Nowadays, with the implemented SCADA system and the installed PMU proliferation, it is possible to define some operation points of the grid. However, the recognition of a global network is an intangible aspect due to the incapacity of affording vast metering installation costs and naturally, the admission of measurements communication faults to systems operators. Such limitations force the presence of auxiliary operations, specifically, the most common is State Estimation that follows the network operation problem to present days. Nevertheless - this is not the only one - the topology processor emerges as an auxiliary function to the system operator control, and actually, that can be crucial to implement a self-healing network, handling all the four mentioned control clusters.

The topology processor function can play an essential role in the control of the network, giving awareness about the real configuration instantaneously. For example, when a circuit fault affects one line, it is necessary a secure reconfiguration of the network, trying to avoid the power supply cut. Such problem can be solved on two stages: the problem recognition and control actions. Firstly, it is imperative to recognise the actual configuration of the network. This auxiliary function acts, providing crucial information to solve the problem efficiently on an unknown part of the grid without system operator intervention.

The definition of a topology processor is not a trivial task, since it interferes with data acquisition to feed such models. The ability to cope with possible incorporated errors defines one of the concerns as well as the lack of observability system. The present dissertation will explore the definition of a topology processor whilst trying to answer to the associated problems.

1.2 Objectives

As mentioned before, the operation of global networks changed significantly in the past decades due to the vulnerability of the events on the grid. The automatism of network operations emerges to increase the efficient response of some unexpected events on the network, avoiding dramatic situations, for example, the energy not supplied and the damaged equipment caused by circuit faults.

In order to take proper control actions, the recognition of the network and states of each critical point is a mandatory idea. Several functions emerge to provide this acknowledgement of the grid, making them a study priority. These functions try to adapt the way that the network operates to a different behaviour. This dissertation focuses on one of those functions - topology processing of the grid. The developed work will accomplish the following objectives:

- The definition of a topology processor based on a CNN capable of accurately determining the connectivity of a line in a certain point of the grid. Such functionality can supply another main operation like State Estimation execution as well as a correct information of the grid. This information defines a correct control action on an automatic paradigm as the self-healing network;

- The evolution of the information areas of the network based on how much a variable is linked to the breaker status - open or close - trying to achieve a better acknowledgement for topology processing operation;

- The definition of a strategy to cope with substation topology determination even if it handles a large amount of lines reconfigurations;

- Planning and optimisation of the Phasor Measurement Unit – PMU – location in order to increase the information areas with the introduction of supplementary information as voltage measurements. Thus, the breaker status classification can be improved taking into account the substation topology concern.

1.3 Dissertation Organisation

The present section emerges from necessity to explain the document organisation composed of six chapters. The first chapter introduces the global concern of operation transmission and distribution networks, the requirement of TSO and DSO to identify the real state of the network inherent variables and the importance of a topology processor as an essential auxiliary tool.

After this introduction, chapter 2 allows an overview of the state estimation problem and how the topology of the network determines the acquisition of correct information. This first part of the chapter also provides a background of the topology determination concern from the first definitions linked to the state estimation problem execution to more recent trends with the application of Deep Learning frameworks with a focus on topology determination only. On the second part of this chapter, an analysis of Deep Learning paradigm evolution is made with a focus on Convolutional Neural Networks's essential proprieties that can contribute to a new topology processor definition.

Chapter 3 develops a different analysis of available measurements based on ITL – Information Theory Learning – where is proposed a new method to define how much a power flow measure is linked to breaker connectivity. Due to this idea, a ranking of power flow results could be done renouncing traditional proximity levels concept associated to the breaker location.

The chapter number 4 is addressed to work developed methodology explanation where all important considerations were done guiding the reader to understand the main problems of training a Deep Learning framework as a CNN. The proprieties of such tool that make a strong candidate to establish as a topology processor are mentioned and how the structure of it can be achieved as well as training procedure of it. Finally, the substation topology problem can be found as the second topology focus problem where a detailed description of this concern is presented.

Chapter 5 proves CNN as a properly topology processor with corroboration of defined ideas on the previous chapter, combined with probabilistic interpretation of data that ITL technique provides. Firstly a validation of proposal CNN models and key configurable aspects is done, followed by proving the influence of an input arrangement, guiding the reader to understand how to attain a desirable topology processor application on a realistic operation scenario. Substation topology determination is mentioned next with the study of two different substations over an information reduction. Finally, the possibility of study the PMU introduction on a group of available measurements is showed with the introduction of voltage content can be found with the demonstration of optimal localisation of it, in order to improve substation topology determination efficiency.

Lastly, chapter 6 is addressed to conclusions of demonstrated work with a definition of original contributions that it introduces on topology estimation concern. The second part of this chapter presents the suggested future work in order to improve the topology processor, based on a CNN. Principally how it can achieve a real-time application as well as an essential tool for planning the metering installation on a study network to an enhance information areas optimisation.

Chapter 2

State of the art

The present chapter provides a theoretical background that allows the comprehension of topology identification concern with some of the essential concepts. The first section defines the state estimation problem where it is possible to understand the necessity of recognising actual states of a power network trying to map it. Following this, a review of existent topology estimators is made with an emphasis in observed problems as computational effort and estimation precision of the actual state of a breaker.

Furthermore, it is crucial to bring awareness about the used framework - CNN (Convolutional Neural Networks) - as topology estimator with a review of *Deep Learning* used techniques, explanation of CNN behaviour and necessary proprieties that make a useful framework as topology estimator.

2.1 State Estimation

This section of the literature review will define the state estimation problem since their foundation by Schweppe *et al.*, first observed issues and mostly the main changes that occurred until the present moment. Besides, the topology processor is mentioned as an essential auxiliary function to state estimation execution. This same function is the central concern of the present dissertation, and this section will provide clarification of past and current proposal methods to solve the referred problem and at the same time what can be done beyond existing models.

2.1.1 Problem overview

Nowadays, for the Transmission System Operator - TSO - and Distribution System Operator - DSO - to operate the energy grid and keeping normal conditions of security and reliability, they are supported by SCADA - Supervisory Control And Data Acquisition. With the same objectives, also occurs with EMS - Energy Management System - acting coordinately with the purpose of monitoring and know the real state of each point of the grid. To help in this task, SCADA and EMS provide the essential tools to operate State Estimation as available measurements and specific telemetry based data, allowing the states mapping of the grid without local direct measurement.

Sometimes, information acquired by SCADA has errors incorporated on it or, in a specific point of the grid, communication of individual measurements fails. This way, State Estimation provides the ability to estimate an operation point of the system with the highest probability possible with the available network results and avoiding to infer false states of the system. In both situations, the central point is preventing TSO and DSO from working with erroneous measurement or fails to take control actions without real knowledge of the network avoiding to cause real damage based on such incorrect actions.

Mathematically, the state estimation problem is defined by an optimisation problem with the following formulation:

$$min\, J(x) \tag{2.1}$$

with

$$c(x) = 0 \wedge g(x) \leq 0 \tag{2.2}$$

where:

- **x** - is the vector of state variables;
- **J(x)** - is the objective function composed by the minimisation of the error;
- **c(x)** - equality-constraint vector;
- **g(x)** - inequality-constraint vector;

Definition of J(x) function represents the errors that will be minimised between estimated values and measured values. In another point of view, considering measured variables equals to real variables plus an error e_i, so that state estimation problem can be defined by [2]:

$$z_i = h_i(x) + e_i \qquad i = \{1, ..., m\} \tag{2.3}$$

$$J(x) = f(z - h(x)) \tag{2.4}$$

with:

- z_i - i^{th} measurement contained bus voltage, line power flows and power injections;
- h_i - i^{th} non-linear correlation between h function with measurements of the state variables x;
- e_i - i^{th} error measurement;

- **x** - vector of n state variables, with bus voltage magnitudes and phase angles;
- **m** - number of measurements;

Alternative State Estimation problem formulation 2.3, is defined by m measurements and n state variables, where $n < m$, imposing more nonlinear functions h_i than state variables x_i to estimation problem. Thus, allows the caracterisation of estimated state vector $\hat{x}$, that will produce estimated measurements $\hat{z} = h(\hat{x})$. Difference between measured values z and estimated values $\hat{z}$ represents, the also known, residual r:

$$r = z - \hat{z} = z - h(\hat{x}) \tag{2.5}$$

While the real value of the estimation problem x is an unknown state, the error e is obtained by the evaluation of residual, r expressed in equation 2.5. Although this is an approximated state, it will enable the resolution of State Estimation problem, as:

$$e_i = z_i - z_i^{true} = z_i - h_i(x^{true}) \tag{2.6}$$

2.1.2 Minimise Errors - WLS

On first State Estimation proposal [3–5], desired optimisation was in charge the minimisation of square error with a matrix R, which provides a weighted optimisation between measured and estimation variables defined in the previous section 2.1.1. That method is called Weighted Least Squares - WLS - and remains one of the most efficient and used tools in State Estimation problem. Hence, it is formulated by:

$$J(X) = [z - h(x)]^T R^{-1} [z - h(x)] \tag{2.7}$$

with

$$c(x) = 0\,, \qquad f(x) \leq 0 \tag{2.8}$$

WLS estimator defines R as a square weight matrix with m dimensions that represent covariance of the errors, so defining $R = Cov(e) = E[e \cdot e^t] = diag\{\sigma^1, ..., \sigma^m\}$ [6], and representing measurement errors independence and distribution error σ_i associated to each element i, with $E(e_i) = 0$ assumption. The diagonal matrix R^{-1} is usually named as $W = diag\{\frac{1}{\sigma^1}, ..., \frac{1}{\sigma^n}\}$, in which each element of W gives information about the reliability of variable x_i. Newton-Raphson is the traditional propose method to solve the optimisation problem defined in 2.7, with minimisation of $J(x)$. Also, problem resolution relies on the first-order differential equality, which is determined by the following condition [6]:

$$g(x) = \frac{dJ(x)}{dx} = -H^T(x)R^{-1}[z - h(x)] = 0 \tag{2.9}$$

Where Jacobian matrix, $H(x)$, of $m \times n$ is:

$$H(x) = \begin{bmatrix} \frac{dh_1(x)}{dx_1} & \dots & \frac{dh_1(x)}{dx_n} \\ \dots & \dots & \dots \\ \frac{dh_m(x)}{dx_1} & \dots & \frac{dh_1(m)}{dx_n} \end{bmatrix} \tag{2.10}$$

Using the Taylor series fundamentals, it is possible to rewrite the non-linear function $g(x)$ in the vicinity of x^k, according k terms, as:

$$g(x) = g(x^k) + G(x^k)(x - x^k) + ... = 0 \tag{2.11}$$

Without losing method accuracy and ignoring the higher order terms of series Taylor resolution to $g(x)$, an iterative solution simplification is given, following Gauss-Newton method [6], by:

$$x^{k+1} = x^k - \left[G(x^k)\right]^{-1} \cdot g(x^k) \tag{2.12}$$

On the other hand,

$$G(x^k)\Delta x^{k+1} = g(x^k) = \bigtriangledown J(x^k) = H^T(x)R^{-1}\left[z - h(x^k)\right] \tag{2.13}$$

with

$$x^{k+1} = x^k + \Delta x^{k+1} \tag{2.14}$$

Gain matrix, $G(x)$, defined as $\frac{dg(x^k)}{dx}$, is a sparse, positive, definite symmetrical matrix. Also, it allows a fully observable system where a set of available measurements are enough to determine a unique state estimation solution without adding more vales required to system resolution problem. Following WLS state estimator, iterative problem convergence is acquired when k iterations are achieved or when a stop condition criterion, generally defined by $\varepsilon > \Delta x^k$, is true.

Resolution of State Estimation problem, as demonstrated previously assumes that necessary data is available or used values have small deviations of real values. Such deviation does not result in a wrong mapping network, and State Estimation formulation converges to desirable results. Regularly, TSO and DSO face other problems like transmission missing measurements - **a lack of observability** - or gross errors incorporated in data acquisition - **false determination of undiscovered states** - both with high potential of a difficult correct decision of unknown variables. Moreover, the system operator has to be aimed by other functionalities, and such tools should support the resolution of the estimation problem. Firstly, providing unknown points of grid like topology configuration and secondly problem resolution with error processing. Some auxiliary proposal functionalities are:

- Bad data processing;
- Topology processor;
- System observability;
- Errors and parameter processor.

Thereby, in this dissertation, the topology processing problem will be clarified with what has been done already and what can be achieved with new approaches, going more rooted in the machine learning field. First of all, a topology problem overview will be presented in the following subsection.

2.1.3 Topology Processing Problem - Classical problem overview

Operating the power grid is a challenging task, as it is always necessary to know the mapping of each power network in a wide area to take properly coordinated operations or apply supplementary recognising action like State Estimation. That emerges with the purpose of network awareness enrichment, mainly where metering tools cannot achieve adequate measurements. Thus, topology processing tools are one of the essential functions to reach such level of an acknowledgement as a result of failure data acquisition by actual SCADA systems incorporated in TSO or DSO. Sometimes, metering equipment in transmission lines fails, transmitting wrong measurements and the loss of communication between them sometimes occurs. Thus, it is vital to deal with possible metering failure to correct topology processing, wherein will feed state estimator.

Thus, there are two clusters of data, analog data referent to power injections, line power flows and voltage magnitudes, contrarily, logical data represents another cluster of data available. Such information is correlated to lines and transformers status expressed for switchers mode, ON or OFF. In the first definition of State Estimation [3–5], Schweppe remarks the impact of erroneous measurements, the analog data, on convergence to a desired solution of the method, if measurements differ significantly of real values, the state estimator cannot ignore them, so infers wrong estimations.

Schweppe *et al.*, brought the Largest Normalized Residual Test - LNRT - as a first technique to deal with bad measurements built on processed residuals 2.5. Being that introduction sufficient to eliminate inconvenient values, without observable system degradation, and WLS estimator will perform next, while the most significant residuals are present. However, LNRT has some issues, beyond more considerable computational effort, was proved when multiple bad data occur like the incorrect interconnection between buses assumptions, the algorithm of residual calculation/elimination performs a wrong bat data identification.

The first developed topology processor [7] with a starter hypothesis test comprehension of the problem, given a degree of certainty to a line configuration. In which if measurements are not sufficient to residual test validation, bad data is the possible answer and the hypothesis of a line configuration is different from the previous topology definition have to be accepted. Mostly of initial works [7–10] was based on residuals test analysis, hence requires a previous state estimation resolution, such methods are classified as *a posteriori* because of that idea. This formulation presents significant doubts where the topology errors cannot be detectable in a critical bus/branch, namely when the most significant residual errors occur in it the system turns unobserved. Such a procedure causes an enormous computational effort as a result of requires the state estimation resolution before each residual test evaluation, in the broadest power network that can mean a lot of residuals reviews. Furthermore, it was proved that topology errors have more impact on state estimation failure than small measure deviations [9]. That fact defines the robustness of a state estimator method, where the ability to cope with gross errors is the essential key to a success estimation.

The evaluation of topology errors existence also can be done before State Estimation resolution, usually defined as *a priori* approach. Thus, Irving and Sterling *et al.* in [11] proposal, a local substation topology analysis based on a constrained minimum-error-modulus solution to achieve a correct switchers configuration, and offers one of the firsts topology processors implementation. Also representing *a priori* technique approach with a practice application, in [12] is defined as a rule-based methodology where it is necessary a sufficient number of rules for or against breaker status corroboration. Otherwise, an undefined identification status occurs with an unclassified breaker.

At the same time, A. Monticelli *et al.* introduce a different perspective to analyse power network topology introducing breaker status [13] representation in estimation problem. Previously, the resolution of zero impedance branches representation with small impedance's approximation was solved by Monticelli defining lines short circuits with state estimation matrix and constrained reformulation [14]. Such contribution shown very helpful to incorporate breakers status in state estimation formulation [13], being a closed breaker represents as a short-circuit branch, differently an open state represents a line without power flow, being this implementation advantageous to take care of a topology formulation. Later, Monticelli defines a new concept [15], Generalised State Estimation - GSE - taking advantage of local observability, wrong data analysis and state estimation common properties in real-time modelling of power networks. Contemporaneously, with some concern about the model effort analysis, a new algorithm appears made by American electric power service corporation, defines a new local strategy based on a local connectivity update breaker status [16].

In a topology point of view, Monticelli brings an initial proposal of WLS estimator to identify suspected zones with most considerable residuals magnitude. When it happened, signify possible recent breakers changes in the first stage of algorithm execution and flag it to a second local analysis instead of deal with all power network breakers. That approach gave a substantial computational complexity reduction comparing with previous topology estimators. All Monticelli works is compiled in the final of 20^{th} century in [2] where GSE is meticulously described. Besides Monticelli method, Ali Abur suggests a similar two stages state estimation procedure established the Least Absolute Value - LAV - estimator [17] trying to avoid the WLS issues. Another example of the application of LAV estimator emerges in [18] where variables were introduced in each branch admittance correlation. Such variables assume three possible values, 1 if there is a connection between two terminal buses, 0 if not and 0.5 if an uncertain status occurs.

Monticelli paradigm influences newest research in this field improving the first stage estimation attempting to prevent WLS estimator incorrect results, surge Abur LAV formulation [17] or Huber M-estimator procedure [19], aspiring a robustness technique minimising gross errors occurrence. Beyond *a priori* research incidence, also second stage of State Estimation suffer some study focus by Clements incorporating the use of normalised Lagrange multipliers for topology error identification [20] as an extension of normalised residuals method in which erroneous circuit breakers, modelled as constraints, and analog measurements could be identified, defining a correct breaker status value.

J.Pereira and V. Miranda developed another state estimation perspective called Fuzzy State Estimation - FSE - with a unique probabilistic data treat measuring the uncertain in available measurements. This new approach is summarised in [21–23] publishes and deeply analysed in J. Pereira PhD thesis [24]. Such approach keeps the binary nature of the topology set as ON/OFF (1 or 0) instead of an interval state where topology variable assumes a value belongs to $[0,1]$. To afford that, variable topology solutions were forced to an $x^2 - x = 0$ function representation [22]. Topology estimation was proved in small power networks with fined adjust of weights in State Estimation formulation enabling real application in DMS and EMS daily operation.

After, based on Mili hypothesis testing identification method [25] and GSE [15] purpose, a probabilistic path guided E. Lourenço, A. Costa and A. Clements to search topology errors even it was critical measurements without renounce the system observability and taking advantage of Bayesian-based hypothesis tests [26]. Same researchers group also develop a method of errors identification involving Lagrange multipliers directly associated with breakers status constraints of State Estimation problem [27] based on Clements Lagrange multipliers introduction [20]. As Monticelli paradigm, the problem resolution was divided into two stages, first of all, collinearity test was used to flag wrong status to a second procedure where the same experiment was performed only to wrong breaker status determinations. Hence, the computational execution reduces with a second stage reduced number of breakers procedure run and the simplicity of collinearity tests.

A. Conejo and E. Caro introduced the latest significant contribution in topology determination field with a quadratic programming optimisation problem definition in a DC approximation of the power network [28]. Identification of erroneous states was provided with the addition of a more extensive set of data in which that fact introduces an extra computational efficiency effort. Due to this issue, that remains the main problem of the newest topology models based on mathematical formulations.

2.1.4 Topology Processing Problem - new paradigm

The traditional methods of topology processing in most of the situations were linked to state estimation mathematical formulation mentioned at the beginning of this section or similar to it. One of the critical problems associated with this procedure is the computational effort, matrix operations of the optimisation problem were required and introduced high model run times. Firstly in 90's a unique perspective presented by Silva *et al.* with the introduction of Artificial Neural Networks - ANN - renounces the also known as the conventional techniques to implement a topology processor based on power network possible states, ON or OFF.

Original works that remark this new approach [29–31] take advantage of feedforward ANN in a supervised training to incorporate a broad set of variables so that the network topology will able to be recognised. Introduction of ANN as topology processor remarks a pattern analysis application with two main contributions:

- Independence between State Estimator and Topology Estimator, founded in multiple offline ANN training input and output network background;
- The capacity of dealing with bad data, correcting them and also critical measurements treatment, eliminating the system observability concern;

Separation of State Estimation problem of Topology determination also disconnect them of significant issues of the initial estimation as system observability checking and offer a bad data reliable method. Such innovation contributes to an advantaged run time reduction and an efficient Topology Processor comparing to traditional ideas.

Inspired by this paradigm Kumar *et al.* try to apply ANN structures as Functional Link Networks (FLNs), Counterpropagation Networks (CPNs) and Hopfield networks to static state estimation and topology processor definition [32]. Hence, CPNs reveals an efficient method to faster topology estimation even with the addition of non-Gaussian noise and corrupt data as input of the stipulated model, giving a step forward in immediate incorporation of them as practical function in EMS or DMS.

Newest architectures of ANNs emerge, in 2013 Jakov Opara *et al.* and Vladimiro Miranda *et al.* suggest the application of auto-associative neural networks or, also known as autoencoders (AE), as competitive topology processors structures [33]. Such work proves the definition of breaker status as a dependency of electric variables, in which local and also neighboured measurements will define the open or close state giving a specific identity to each one [34]. So in work developed by this investigators [33–37], was established a different perspective of decentralised AE application where each breaker was defined by a specific competitive AE structure based on historical data set variables dependency, creating a *mosaic of estimators*. In later works, unsupervised AE training techniques and correctly selection of variables that accurately define the state of breaker was tested to improve efficient test cases.

AE incorporation in power networks topology estimation takes an essential step in establishing an auxiliary tool to state estimation execution for the reason that was achieved a faster tool with a mosaic of estimators structure. Other outstanding contributions of the suggested mosaic idea are computational effort reduction of topology processor execution and the possible application of deep learning frameworks like AE to several cases as extensive power networks and as well as complex substation internal topologies. All this work results in PhD Jakov Opara thesis where is possible found detailed explanations about progress in topology processor investigation AE framework [38].

2.2 Deep Learning - an uncommon approach

The present section emerges on the necessity of explaining *Deep Learning* as a common area on scientific and engineering field, the essential foundations beyond that, common frameworks and properties that change the way of investigation, demonstrating a new approach on achieves better results than the traditional path of made topology processors.

2.2.1 Deep Learning vs Classical methodologies

Since the appearing of the computer, the human being also desired that that box machine could resolve all the problems. Such desire follow the next generations to accomplish the newest problem resolutions like simply solving mathematics expressions automatically, voice identification or different types of image recognition. Nowadays, fast and precise answers to several issues are required to take actions on adequate resolution time. The most rapid evolution of technology cause this necessity, an increase of quantity information available induced by the exponential development of metering tools such smartphones with high-resolution cameras and microphones, telemetry sensors of high scale resolution or any collecting data toll, enables large data repositories.

The main problem associated with this amount of data was that classical mathematics models could not handle with such amount of information, because program run time increase and the efficiency resolution remains unaffected. Thus, Deep Learning offers a different procedure idea of processing data, based on simple mathematical rules and recognition features on available input, mapping that in global features necessary to attain desired output or the most accurate one. Scheme 2.1 focuses on the key steps between input data and desirable output comparing rule-based systems. The machine learning approaches adds features concept incorporated on input and further deep learning where simple features are collected to abstract levels of information with sufficient layers to achieve representative features to produce a proper answer to the problem that such an algorithm is affected too.

Hence, complexity models perhaps require non-intuitive structures to common human perception but also are usually based on biological neural action with non-linearity incorporated on it. Actual models where Deep Learning is implemented can classify untagged data like images with a low resolution where human vision is not capable of identifying that. However, unseen data also can be converted to identify information from these tools since an important data set is available 2.2. Although it is a crucial distinction between the quantity of information and representative

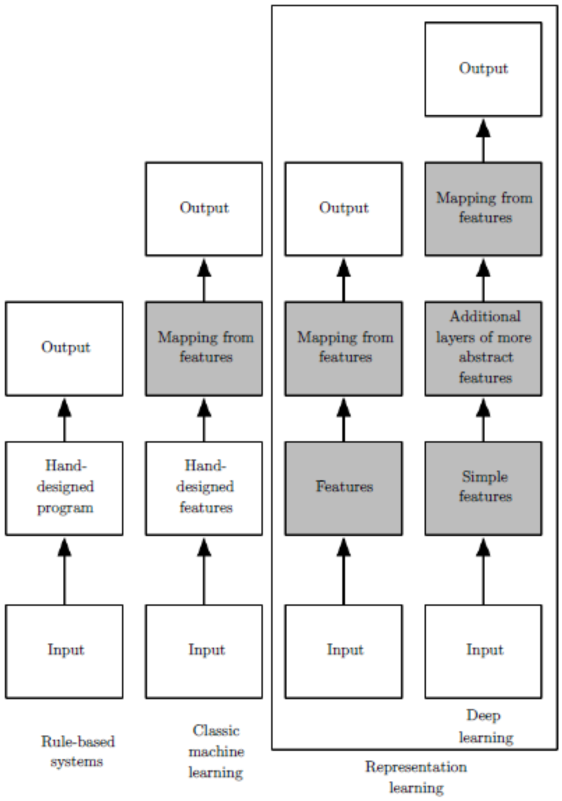

Figure 2.1: Flowcharts representing differences between traditional programming approaches, classical machine learning and, what can it is possible to achieve in the AI field based on machine learning techniques [39].

cases present on input data, the most extensive set of values perhaps not represent a variable data set. On a dataset, if some cases are replicated will not contribute to the training procedure of a framework, and if the newest input appears, a targeted output probably will not occur with desirable accuracy. So, it is essentially assured that where a data set is generated the most critical characteristic on it is the representativity of cases, sometimes also enforce a big data set, depend on the complexity of the problem, but is not the main rule.

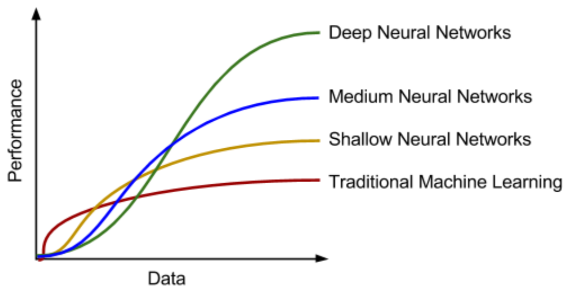

Figure 2.2: Evolution of performance on different common applied types of machine learning with the amount of available data [40].

Figure 2.2 reflects the capacity of Deep Neural Networks to attain uncovered features and reach high-level performances with larger representativity of data representing all possible problem samples. Another developed Machine Learning techniques also improve their performance with the increase of available information but less than the first one. What is essential understand is that with Deep Learning the increment of data and representative cases allow better performances than classical models in which uncover features could not be observed and consequently their run time, as well as the accuracy of the model, would persist the same.

During past decades some *Deep Learning* frameworks were developed as a result of different complex problems applications. That same fact returns a gain of popularity on the engineering field with essential applications. In this dissertation, will be possible the application of one framework, CNN, as features collector in classify topology problem. In the following subsection, this framework will be clarified with a focus on principal proprieties to attain better performance than existence topology processors.

2.2.2 CNNs - Convolutional Neuronal Networks

The importance of visual pattern recognition toke firsts steep in 1989, where LeChun *et al.* introduce digital classification problem [41] giving dataset with 2D images digits in a range between of 0-9 has the main goal of classifying them. In history, this is the most straightforward problem and at the same time, the base problem to corroborate a new suggested algorithm. Following this problem, LeChun in [42] established the firsts efficient Convolutional Neural Network - CNN - implementation to classify correctly digit numbers.

This cluster of neural networks is based on the animal visual cortex and the spatial correlation between neurons, wherein each neuron is subjected to stimulus, and it responds like a convolution operator trying to attain patterns on visual input. Thus, CNN has the principal characteristic the grid-like topology, where a convolution operation is applied to a 2D input, in which is simply a pixels matrix where each pixel represents a specific number, and the spatial identity is conserved. This type of neural networks is a particular type of Feed Forward Neural Networks group with the principal difference that CNN architecture can use fewer freedom parameters, weights (w) and biases (b), than Feed Forward Networks. That fact did LeChun thought CNN as an essential classification framework because it can deal with a considerable amount of input elements where an image with many pixels is a complex and desirable target.

In the following sections, the structure of CNN will be explained, the unique proprieties that CNN has compared to regulars ANN making CNN a framework to taking into account as topology processor. Finally, a historical overview is presented to understand the problems that such tool faced and if breaker status reconstruction is a possible target of CNN.

2.2.2.1 CNN - Architecture structure

In regards to architecture that this type of neural networks can exhibit, each layer is characterised by one of two primary operations, convolution or polling, representing the convolutional layer and pooling layer. Also, CNN frequently has a fully-connected layer, generally at the end of the neural network like regular type of neural networks, connecting it to outputs neurons, giving a correlation between CNN outputs and the desired targets. The arrangement of such layers are associated with the dimension of the input image, and it is possible to see a vast diversity of constructions linked to a particular problem.

Fundamentally on a fully-connected layer, it is crucial to understand that neuron interpretation stands similar to previous neural networks, where x_i inputs are affected by certain degrees of freedom, w_i and b_i 2.3. All these contributions on neuron body are subjected to a non-linear function interpretation trying to bring out neurons real proprieties. Such approximated behaviour is usually called an activation function, and the most popular function is a hyperbolic tangent function (tanh(.)), but also sigmoid and ReLU (Rectified Linear Units) functions are used and in most situations to perform a better result in classification problem 2.4.

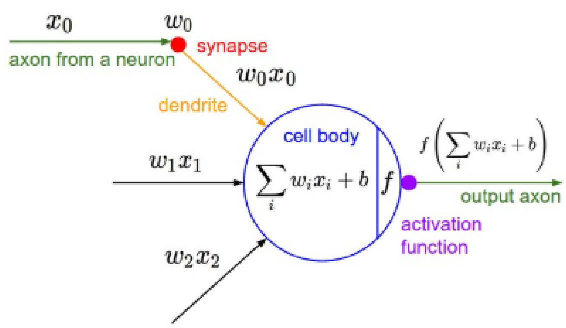

Figure 2.3: Approximation of neuron comportment to apply in neural networks with weights w_i, biases b_i and input x_i [43].

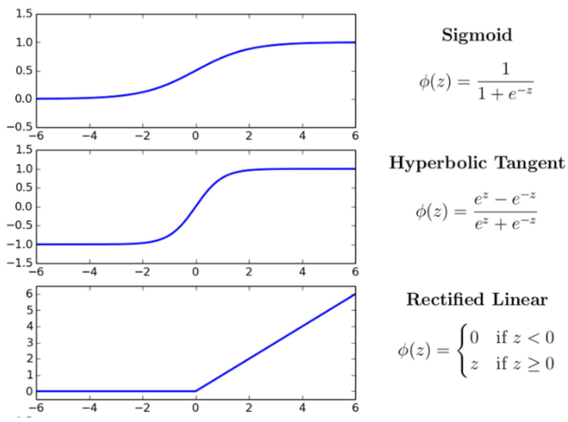

Figure 2.4: The most used activation functions in neural networks field: Sigmoid (ON TOP), Hyperbolic Tangent (MIDDLE) and ReLU (DOWN) where the input z is affected by them.

What makes CNN a particular type of neural networks is the convolution operation that is conducted by three main characteristics: sparse interactions, parameter sharing and equivariant representation of features [39]. In a convolution operation 2.5, a steady matrix of weights, also known as the kernel, have a 2D dimension smaller than the input image and giving the referred sparse connectivity feature in which just the most significant neurons contribute to second layer interpretation. Also, in figure 2.5, it is possible to see that same kernel inspects all the input data to find features that representative CNN with training, will be able to detect. Well, the classification problem depends on such features that will find on input images essential identity points. Thus, in opposition to most of the neural networks, the fact that kernels/filters remain the same - parameter sharing - will reduce the computational effort when given a more significant number of input data and, instead of small image pixels, like digits problem recognition, would be more attractive bigger images and a large amount of applications CNN will perform.

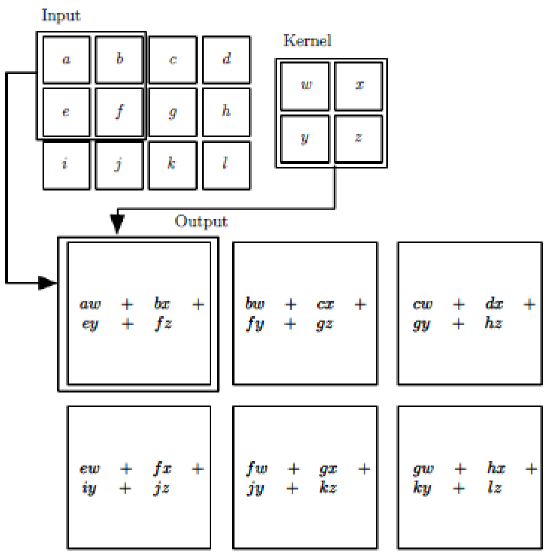

Figure 2.5: Visual example of a convolution operation with a clear demonstration of sparse interactions, parameter sharing and equivariant representation proprieties with a matrix of weights (kernel), and input image (pixels matrix value representation) [39].

Always associated with a convolutional layer is a pooling layer, the principal responsible for dimension reduction of features maps earlier obtained by convolutional layer. That layer has a crucial function in the extraction of the maximum value of convolution results. Usually, this type of operators is made of filters, with dimensions depending on input image size, travelling along with the previous filtered values, step by step, extracting the maximum incident value. In other words, the maximum extraction of value means the most representative characteristic presenting in the selection area and clustering them. Such downsampling of input run until having a trade-off between minimum dimensions and information content preserved.

Thus, a fully-connected layer is applied to connect the last iteration to output neurons, where all previous neurons are flatted in a 1D vector and connected to output neuron. In fact, in most of the cases, two or three fully-connected layers can be found to prevent a drastically dimension reduction. The main function of this last layer gives a probabilistic interpretation to results of convolution/pooling operations, where the maximum value - desirable output classification - the result of a *softmax(.)* operation 2.15. Indeed, with $z = (z_1, ..., z_k) \in \Re^k$, *softmax(.)* the function can be represented by a normalised exponential function where the result is comprehending in $[0, 1]$ range:

$$softmax(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad j = 1, ..., K \tag{2.15}$$

However, the CNN architecture structure can assume different forms, and it always depends on the initial complexity problem. In a digit problem recognition 2.6, two levels of convolutional and pooling layers would be enough due smallest 32x32 input image, yet, a problem with the most substantial amount of unique features may need more operations layers too. Besides that, CNN will always preserve the capacity of transforming local features into high-level maps with global features strictly necessary for a quick classification resolution.

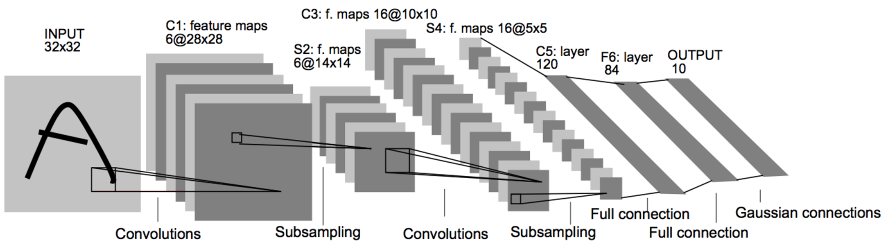

Figure 2.6: Demonstration of an example of CNN architecture, LeNet-5 used to digit classification where it is observed the three principal layers: convolutional, polling and fully-connected layer [42].

2.2.2.2 CNN - Past and Present

As mentioned, LeChun *et al.* demonstrate the first CNN application in digital recognition problem [42] but also Steve Lawrence at same time purpose a hybrid method of local image sample representation, a SOM - self-organising map - network and a convolutional network to face recognition [44]. Forward, with technology development in the new century, Deep Learning area also reflects the same growth and CNN was sawed as a tool to face new challenges in classification obstacle. Major companies such as Google and Facebook, take larges steps in face identification with CNN as a framework [45, 46].

In image recognition problem, CNN suffers a significant evolution and proves it is valued as an efficient tool. Indeed, it encourages the application of that on large-scale video classification, where the variable time takes into account as an essential problem. In [47], time consideration was solved, giving clips with fixed frames in order to not affect CNN architecture structure. This problem was interpreted as a 3D problem and can be found in [48–50], a new approach that guides the CNN structure development interpreting the input as a volume where the time dimension is preserved, and newest problems were addressed like human action recognition and real-time video classification.

Study of CNN brought R-CNN (region-based convolutional neural networks) structures [51, 52] where regions with specific features are detected in a multi-size image input and massive application of DCNN (Deep Convolutional Neural Networks) [53,54] with dense features analyses and decomposition. All these variations add run times decreasing, intricate structures in which enable new problems approach changing the CNN state of the art.

Nowadays, CNN and variations of them, are efficient tools in image classification and pattern recognition, so implementation of these ideas in new areas is a paradigm evolution. Recently, the biomedical field has fruitful experiences in tomography images classification, and biological data processing using CNN and composite structures with autoencoders and deep-belief neural networks [55, 56] providing an auxiliary and powerfully tool detecting hidden irregularities as unseen features.

In Power Systems field, recently CNN has seen the first application to detect dynamic events in power networks as the generator and line tripping also as load disconnection, and inter-area oscillation based on frequency variation in time dimension captured by phasor measurement units - PMU - [57]. In which, was demonstrated that a continuous variable as the frequency could be represented in an input image feeding a CNN model and with a correct representation of them can produce a specific group of essential features characterising each event. That fact proves CNN utilisation on a non-image input and at the same time achieve an efficient performance.

2.3 Final Remarks

The correct representation of power grid was defined in present chapter with particular concern to what is necessary do in order to define a most accuracy topology processor taking into account the challenges inherent to this investigation such as the observability of the system. TSO and DSO handle with a lack of operating networks information as well as the urgency of having the acknowledgement of network configuration in an adequate time to take supplementary control actions based on such acknowledgements like State Estimation execution or control procedures on the network.

Deep learning neural networks prove some significant benefits a long time ago compared to classical models. Each framework with deep-rooted features and so that CNN emerges as a framework with specific proprieties as the possibility to defines a 2D input. That characteristic was taken as an advantage to pattern recognition by input measurements, and a better learning procedure with vicinity establishing correlations to an easy training procedure attaining better performances.

Chapter 3

ITL - Information Theory Learning

This section will provide an approach to the origins of Information Theory and the applicability of this concept to extract information content of variables linked to network operation, considering the topology estimation problem. Throughout the years, engineering judgement criterion was an imperative role-model to classify the importance of power flow results referred to a specific point of the grid, such as a breaker where proximity levels were considered. Demonstration work has the main goal of adding a practical and intuitive tool, in order to classify the most critical variables of the system related to breaker status, connected or disconnected.

3.1 Brief introduction to Information Theory Concept

Around the middle of the XX century, mathematician and electrical engineer Claude Shannon *et al.* encouraged by transmission messages under a noisy-channel problem, and mainly, the author of the reconstruction of a receiver signal with a low probability of error [58]. With the definition of that concept, a quantification of information emerges, encoding on the transmitted signal, and so that entropy appears as the amount of uncertainty incorporated on a set of random values. This is the first attempt to achieve an informative quantification goal.

Work developed by Claude Shannon immediately affects positively other study fields as statistics, cryptography and electrical engineering as well, establishing a new quantification of unknown values and a remarkable relevance of his theory information concept proposal. The next topics refers to important definitions in information quantification, started by Shannon and developed by other respectable mathematicians. Such basics definitions guide present work to quantify some available measurements, for example, the power flow metering variables referent to the breaker connectivity.

3.1.1 Shannon's Entropy

Entropy, as defined by Shannon in the information theory [58], is a measure of uncertainty associated with the system where a variable is inserted and subjected too. Considering a variable X, where $x_i \in \Re^D$ and a random sample A of n pairs $\{x_i, y_i\}$, a probability density function - pdf of

X is interpreted by $P = \{(x_i, p_i(x_i), i = 1...n)\}$ or just $P = \{(x_i, p_i)\}$, the definition of Shannon's Entropy is:

$$H_1(X) = E[-log(P)] = -\sum_{i=1}^{n} p_i \, log \, p_i \tag{3.1}$$

With

$$\sum_{i=1}^{n} p_i = 1, \; p_i \geq 0 \tag{3.2}$$

On this idea, Shannon determines the uncertainty of the system X as the sum of entropy across the x_i values characterised by the probability p_i, the entropy of each element x_i is given by $log \, p_i$ that contributes to the global entropy.

Considering the continuous domain and incorporated variable X, the pdf of X is $p(x)$, so the entropy is defined by:

$$H_1(X) = -\int_{-\infty}^{+\infty} p(x) \, log(p(x)) \, dx \tag{3.3}$$

The Shannon Entropy definition measures the ambiguity of information expressed by X and is valid for both the continuous and discrete domain. On signal transmission's study area, where this issue emerges, the content of information is obtained in bits, so in the previous definition, the logarithm function basis is 2, and so that, affording that meaning. However, the basis of logarithm function can assume various types, according to the behaviour that the application of it requires.

3.1.2 Renyi's Entropy

Mathematician Alfred Renyi starts his work in theory information with the purpose of developing the concept of entropy established by Shannon, in which, Renyi characterises Shannon's entropy as part of some entropy family functions. This way of measuring the uncertainty of the system introduces parameter α, that specifies each entropy function without compromising the objectivity of information content measure.

Considering $P = \{(x_i, p_i)\}$ that defines the random variable X with $x_i \in \Re^D$. Renyi's family functions is established by:

$$H_\alpha(X) = \frac{1}{1-\alpha} log(\sum_{i=1}^{n} p_i^\alpha) \tag{3.4}$$

With

$$\alpha > 0 \wedge \alpha \neq 0 \tag{3.5}$$

Proving that Shannon entropy belongs to this generalisation of family functions is not obvious because when $\alpha = 1$ the expression 3.4 mathematically diverges. Therefore, analysing the neighbourhood of this point, Renyi proves that when $\alpha \to 1^+$ and $\alpha \to 1^-$, Shannon's Entropy can be obtained as:

$$\lim_{\alpha \to 1^+} H_\alpha = \lim_{\alpha \to 1^-} H_\alpha = H_1 \tag{3.6}$$

As demonstrated in the equation 3.6, the Renyi's entropy family functions converge bilaterally to Shannon's definition of entropy 3.1. This cluster of functions, where Shannon's entropy is included, show a gateway to classify an unlimited number of entropy functions that with a parameter β arise:

$$H_\alpha \geq H_1 \geq H_\beta \tag{3.7}$$

With,

$$0 < \alpha < 1 < \beta \tag{3.8}$$

Interpreting the equations 3.1 and 3.4, the main difference between Shannon's and Renyi's entropy measurement remains on logarithm function action when the system's entropy is obtained. So, in definition 3.1 $log\, p_i$ is weighted by the probability mass function p_i, and the logarithm function acts in each element i. Otherwise, in expression 3.4 logarithm function leads with the sum of all p_i distributions power by a factor α.

The computational effort of this entropy measurements deals with the placement of logarithm function. That fact makes Renyi's entropy faster and more efficient than the first formulation of entropy 3.1 that takes in count the sum of the n logarithm functions, which does not happen in Renyi's family functions that apply the logarithm functions just one time to the sum of n p_i^α distributions.

3.1.3 Renyi's Quadratic Entropy - a particular case

A specific case of Renyi's entropy family functions - Renyi's Quadratic Entropy - is defined when $\alpha = 2$. Assuming a random variable X characterised by the probability distribution $P = \{(x_i, p_i)\}$, as

$$H_2(X) = -log(\sum_{i=1}^{n} p_i^2) \tag{3.9}$$

Considering the continuous domain,

$$H_2(X) = -log \int_{-\infty}^{+\infty} p(x)^2 dx \tag{3.10}$$

This particular case takes some importance in a graphical interpretation of entropy. Assuming purpose probability distribution $P = \{(x_i, p_i)$ mentioned in the previous subsection, n defines spacial dimensions where the entropy of the variable X can be allocated in a hyperplane given by:

$$\sum_{i=1}^{n} p_i = 1 \tag{3.11}$$

The Euclidian distance of x_i point in such hyperplane to the origin, in which each axis is a p_i contribution, subject to negative logarithm function provides Renyi's Quadratic Entropy.

3.1.4 Parzen Windows method

Sometimes, in information theory field, when data is analysed to measure information content like entropy, only a discrete set of values is available. This way, in order to explore all potentialities of the available measurements it becomes usefull to know the pdf representation of each value, allowing further exploration with information theory tools.

Histogram representations have been used to describe the uncertainty of a discrete value and give a visual estimated representation that this point can take. However, histogram representation forbids a mathematical and precise representation of the randomly that a set of variables assume. Moreover, combining these variables is a more laborious task, mainly when having a more significant number of samples.

In order to provide an answer to the pdf estimation problem, the American statistician Emanuel Parzen, developed his work based on kernel density estimation [59]. Thus, Emanuel Parzen to estimate the hidden pdf, suggests represent each n points $y_i \in \Re^M$, i=1,...,n in an M-dimensional space by a kernel function centred in each y_i. Such representation helps to measure the influence between y_i samples approximately, by the sum of each contribution, as kernel functions provide statistical content to previous discrete values.

So, defining a Gaussian kernel function, a pdf estimation $\hat{f}$ of the exact f_y distribution is demonstrated by:

$$\hat{f} = \frac{1}{n} \sum_{i=1}^{n} G(z - y_i, \sigma^2 I) \tag{3.12}$$

Emanuel Parzen also proves that when $\sigma \to 0$ and $n \to \infty$, $\hat{f}$ converges to real pdf $f_y(z)$, such improvement helps in method validation. Taking in count a random variable X = [-2.2, -1.6, -1.1, -0.9, -0.2, 0.4, 1, 1.5, 2, 2.9] of 10 discrete elements, an illustrative example of Parzen Windows method:

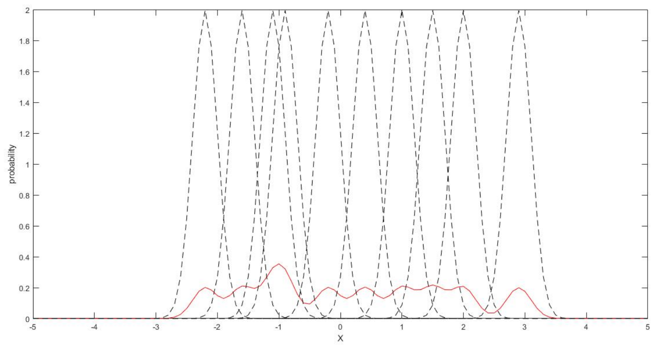

Figure 3.1: Parzen Windows method applied to X with $\sigma = 0.2$. The black dashed Gaussian curves represent pdf of each x_i element and the red curve the pdf estimation following Parzen Windows technique.

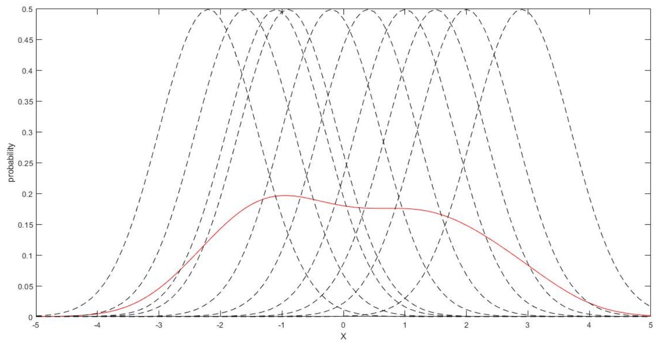

Figure 3.2: Parzen Windows method applied to X with $\sigma = 0.8$. The black dashed Gaussian curves represent pdf of each x_i element and the red curve the pdf estimation following Parzen Windows technique.

Analysing the demonstrative example, the kernel deviation σ takes some importance in size and shape of final pdf estimation $\hat{f}$. A larger size of σ 3.2 provides a soft variation of the pdf curve, while in example 3.1, a small deviation of x_i centre value makes a unique pdf with more singular element observation. In this case 3.1, the influence on the neighbourhood is less than 3.2's example, meaning that using the Parzen Windows method is not intuitive, a previous study of data is necessary to select the right σ, given a desirable representation of data.

3.1.5 Distance of Cauchy-Schwarz

Another important concept that allows the comprehension of developed work in order to measure variable content is the distance between two different pdf, also known as inequality of Cauchy-Schwarz. This new concept introduced by these mathematicians goes back to the beginning of the 20^{th} century with the proof of inequality $|\langle u,v\rangle| \leq \|u\| \, \|v\|$ being, u and v as regular vectors. Such elementary definition is helpful with different application fields as linear algebra, vector algebra and most relevant to the present dissertation problem, with the implementation of probability theory understanding. So, given two continuous random variables X and Y representing by pdf $p_x(X)$ and $p_y(X)$, the distance of Cauchy-Schwarz, D_{CS} is given by:

$$D_{CS}(X,Y) = -log \frac{\int_{-\infty}^{+\infty} p_x(X)p_y(X)}{\sqrt{\int_{-\infty}^{+\infty} p_x^2(X)dx \int_{-\infty}^{+\infty} p_y^2(X)dx)}} \tag{3.13}$$

Defining the distance of Cauchy-Schwarz is also possible to a discrete domain and converge to initial vectors proof representation by this mathematicians. Thus, the equation on 3.13, representation by discrete finite distributions p_x and p_y, upper bounded by n elements in a discrete domain can be represented by:

$$D_{CS}(X,Y) = -log \frac{\sum_{i=1}^{n} p_{xi} p_{yi}}{\sqrt{\sum_{i=1}^{n} p_{xi}^2 \sum_{i=1}^{n} p_{yi}^2}} = -log \frac{P_x \cdot P_y}{\|P_x\| \, \|P_y\|} \tag{3.14}$$

In a discrete domain, it is possible to understand the representation of pdf as spacial vectors with specific characteristics and the distance of Cauchy-Schwarz remains in a proportion between vector multiplication and its norm multiplication 3.14. However, this formulation is delimited by different axioms that give unique proprieties to this definition of probabilistic distance. In which, principals rules verify:

- Symmetrical correlation, $D_{CS}(X,Y) = D_{CS}(Y,X)$;
- min $D_{CS}(X,Y) = 0 \Leftrightarrow P_y = P_x$;
- $0 \leq D_{CS}(X,Y)) \leq \infty$;

Comprehension of this concept is facilitated with an illustrative example, where proprieties referred before, emerges with a practical point of view. Considering three different uniform distributions p(x), z(x) and q(x), with respectively intervals [-2,0], [-1,1] and [0,2] graphically as:

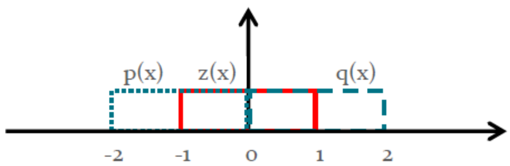

Figure 3.3: Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with p(x) (tiny blue dash), z(x) (red line) and q(x) (strong blue dash) [60].

When pdf did not overlap in their areas, the event represented by p(x) and q(x), the distance between them is calculated using the equation on 3.14, where upper bounded value is obtained

$D_{CS}(p,q) = \infty$. Also, minimum value can be demonstrated where two same distributions were considered, for example, two distribution p(x) as equal distributions define a ratio in 3.14 of value 1, in which, transformed by logarithm function defines a null distance $D_{CS}(p,p) = 0$.

In another hand, symmetrical propriety quickly appears in this example, mainly where z(x) is considered. Observing, $D_{CS}(p,z)$ or $D_{CS}(z,p)$ it is possible to infer that the overlap area is the same, so the result of distance calculation brings a same non-negative value, where it is as high as the distance between them.

3.2 How much a measurement defines a state of a breaker

Since topology concerns started, every model mentioned in section 2.1.3 and 2.1.4 takes into account the importance of measurements that will feed the proposed topology estimator. Most of the topology processors consider all measurements present in the system, which is a theoretically valid idea. However, in a real application scenario, just a few power flow results are available and in that case, it is essential to select the most relevant ones for topology determination.

Using of so-called engineering judgement criterion, it is possible to infer that most important variables to a breaker status determination (ON/OFF) are located near that breaker point. Direct measurements as power flow in line that a breaker is placed and power injection in buses that delimit such line are the most important. The real issue occurs when such direct measurements are not available, and raising such questions as: **Which power flow results choose next? Which are more correlated to breaker status?**

Naturally, the distance to breaker location is an important point to takes into account. A power flow based too far away from the breaker does not contribute to this ON/OFF status, as well as a closest variable of power flow estimation.

Most investigators use engineering judgement criterion to define such content value with levels of proximity, in which, the first level is associated with lines and buses next to a reference point. The second level is next to the first level, and so forth, achieving levels much far than the previous one with less attributed importance to far away measurements. **This approach gives a global idea of the system, but that occurs in fact? Do all the power flow results at the same level have similar weight in breaker point of view? Can a measurement located in a second level be more important that other in a first level?**

The first relevant work that meant to give answers to this concern and also inspired this dissertation's main area of study was made by Jakov Opara [35] who used the concept of mutual information to rank the most important power flow results to supply autoencoder models, the topology processors mentioned in previous section 2.1.4.

In this chapter, a new method of ranking power flow results with more content representation will be clarified, based on the distance of Cauchy-Schwarz between two pdfs. First, a demonstration of the topology problem in a probabilistic point of view is necessary. Given a variable X with a set of discrete values x_i with a *Parzen Windows* technique, it is possible to reconstruct the pdf P(X) in order to represent the global distribution of X in domain x with an adequate deviation σ.

Topology concern consists in a binary problem, meaning the breaker can be connected (ON) or disconnected (OFF). These two existing options define topology variable, Y, allowing the separation of P(X) in two other pdfs where the state is known *a priori* - Y = ON or Y = OFF - giving two essential pdf, P(X|Y=ON) and P(X|Y=OFF), showing the variability of values X as a function of topology state Y.

Considering a considerable amount of data scenario and an approximate equality representation of situations ON and OFF in each x_i value of the variable X, the demonstration of $P(X|Y = ON) \times P(Y = ON)$ and $P(X|Y = OFF) \times P(Y = OFF)$ gives a real mean value, the distance between them can determine how much values of X differ with a binary value of Y. This explanation can be more elucidative with examples figured in 3.4 and 3.5.

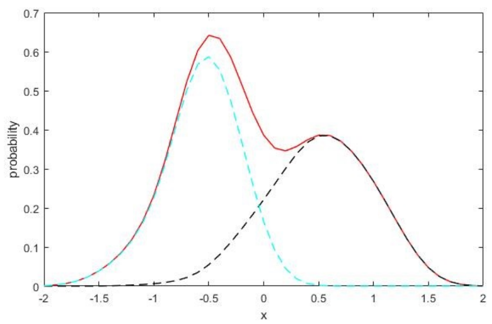

Figure 3.4: Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with P(X) (red line), $P(X|Y = OFF) \times P(Y = OFF)$ (dashed black line) and $P(X|Y = ON) \times P(Y = ON)$ (dashed light blue line)

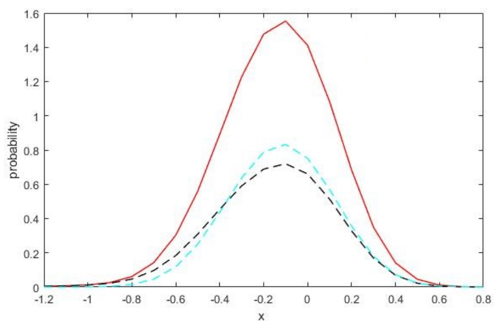

Figure 3.5: Illustrative example of pdf correlation on the distance of Cauchy-Schwarz calculation with P(X) (red line), $P(X|Y = OFF) \times P(Y = OFF)$ (dashed black line) and $P(X|Y = ON) \times P(Y = ON)$ (dashed light blue line)

Differences between these two examples are clear, to the same σ definition of Parzen Windows method reconstruction, in the first figure 3.4 it is possible to recognise the uncertainty of values that a variable takes. When the state Y changes, a range of P(X|Y=ON) in domain x is distinct comparing with range of P(X|Y=OFF), producing a distance between them $D_{CS}(P(X|Y = ON), P(X|Y =$

$OFF)) = 1,54$. A considerable value to take in count when looking at most important variables correlated to a breaker somewhere in the power network.

However, a larger coincident domain region to these two pdf, may not give distinctive information about the breaker connectivity. This coincident region creates a misunderstood area that could take massive proportions and example figured in 3.5 represents that. Both of pdf, P(X|Y=ON) and P(X|Y=OFF), are almost coincident, a value that X can take possible represent a state ON or OFF, naturally the calculated distance results in a null distance $D_{CS}(P(X|Y = ON), P(X|Y = OFF)) \simeq 0$. So, in other words, it may refer that measurement X not characterise the state Y of a breaker.

3.3 Final Remarks

ITL techniques brought a unique interpretation of data in different study fields. In this chapter a more recent methodology was proved in order to systematic ranking measurements of power flow linked to a line breaker. Furthermore, real applications take into account the lack of information, and that is important defines the most appropriate measurements to breaker status reconstruction.

Moreover, the engineering judgement criterion has imprecisions, providing a qualitative classification comparatively to a quantitative definition by D_{CS} information methodology, classifying each power flow variable distinctively and adequately.

Along with this dissertation and work explanation, the developed tool will be mentioned as part of an informative technique to help in choosing variables, and input value organisation of topology estimator that will be introduced later, with more practical examples proving the advantage of taking in count informative techniques.

Chapter 4

Topology processor based on CNN

After contextualisation of fundamental ideas and the main problem of topology processing issue, Deep Learning technique as CNN and ITL concept emerge as the essential areas to the developed work in this dissertation. The present chapter firstly has the main goal of clarify the test system, how the data set was generated and preprocessing that suffers before feeding CNN models. Also, such models are detailed, being the essential concern of this work and how input values could be defined as an important part of accurate results. The training procedure is also mentioned as the principal concern of single breaker connectivity determination, as well as substation topology that is defined in the final section of the present chapter.

4.1 Test system - dataset generation

Before defining the CNN models, a valid data set to perform correspondent developed topology processors was necessary, and the IEEE 24-bus test system was chosen as a proper transmission network example of an extensive and real power system.

Furthermore, all of the neural networks need a representative number of different events, in this case, working points of power network to incorporate in training and test procedure. Hence, it was important to consider different scenarios of operation with variable levels of load and consequently corresponding power generation. According to it, a dataset based on power flow execution with the next properties was accomplished:

- Load level with probability distribution 4.1 in power flow supply and a variation of $\pm 10\%$ from the generated case;
- To power flow results were also added a Gaussian noise with a standard deviation of 0.005 p.u. in 100 MVA power system base to power and voltage magnitude variables;
- Variable topology arraignment of 10 breakers 4.2 considering two possible states of each one, connected or disconnected.

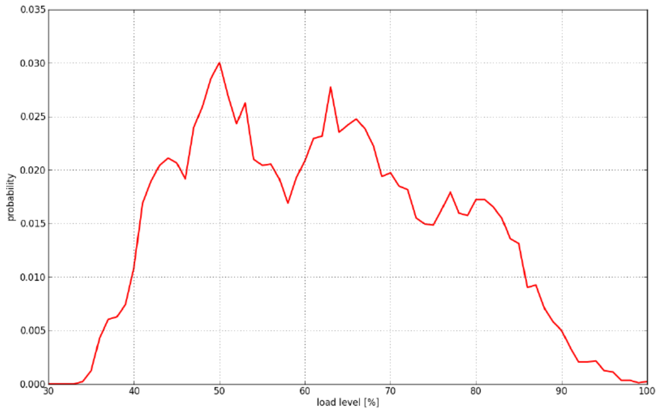

Figure 4.1: Representation of load level as pdf in power flow estimation.

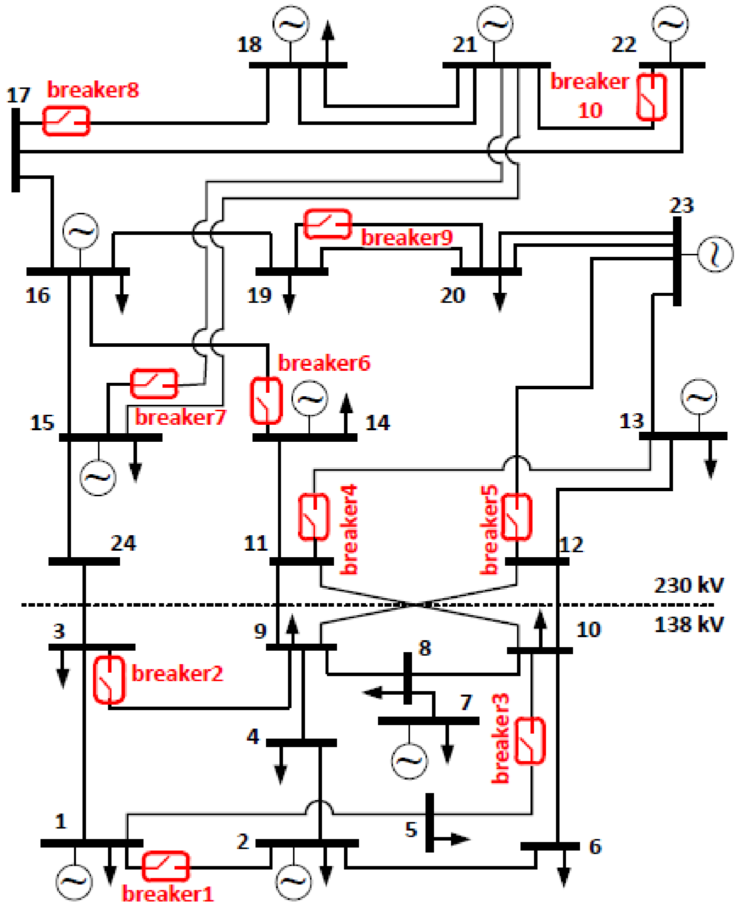

Figure 4.2: Breakers arrangement in the test IEEE RTS 24-bus system.

According to this, 20000 possible scenarios were generated and adjusted, considering previous rules to real case approach characterising the training and test models on adequate data. Also, it is important to refer that, in the dataset, a binary possibility - ON or OFF - of each breaker is approximately represented in half samples to open switch mode and the remaining to close one.

Selected breakers take into account different possible situations: lines delimited by 1 or 2 PV buses, parallel lines and breaker located between PQ buses, all of them distributed along the network 4.2. Thus, a large and representative data set was performed as well as supplementary information about network parameters can be found on appendix A.

After this first step, preprocessing data is required before measurements serve as input to the neural network model. In the resulting dataset, an extensive range of values with distinctive magnitudes can be found, near to zero or much higher than it, such scale of values jeopardise the neural networks learning and free parameters optimisation. As a result of this concern, three types of preprocessing were made, two of them based on normalisation where magnitude was reduced to a range of values $[0,1]$ and $[-1,1]$. The last procedure strategy is called standardisation, meaning that the dataset values were transformed on input values with the application of a Gaussian distribution with zero-mean and unit-variance, trying to preserve the distance between values but also reducing the magnitude of them.

4.2 CNN model

Defining the model of CNN as features collector of a 2D input is not evident, the definition of filters and their quantity does not have a specif rule model. In addition, CNN cannot be built on just convolution and pooling layers, making it important to give real mean to output result of such operations. The proposed model is based on another two necessary operations, Multilayer Perceptron and Logistic Regression, defining a classification tool that will be formulated in the next topics, with corresponding training and test procedure description.

4.2.1 Classification phase

As mentioned before, CNN's principal propriety is pattern recognition in an considered image input. If an image is easily recognisable to the human eye, for example, identifying in an image if it is a person or not, a set of values do not infer a piece of perceptible information about them. Accordingly, CNN with convolution and down-sampling operations, in the end, extracts final values also with non-detectable information as an answer to problem itself.

Therefore, it is essential to define a way to converge this filtered values to a real meaning binary solution. Usually, algorithms affected with CNN layers also use a final hidden layer between them and the classified events, in these case, breaker connected or disconnected, defines the classification procedure based on a particular type of Multilayer Perceptron - MLP.

MLP is a class of feedforward artificial neural networks founded on the principal function of neurons, where outputs are a weighted input affected by a constant bias. Architecturally, it is composed by an input vector and hidden layers until output neurons. The possibility of composing a different number of hidden layers is a fundamental characteristic, in order to approach problems with a large number of input values to fewer outputs. That MLP propriety allows a progressive downsampling of initial values, providing a deeper structure and attaining better results in complex problems establishing a more deep correlation between input and output.

The chosen MLP considers the existence of just one hidden layer, the group of input values was not sufficient to contemplate more than one hidden layer. That fact jeopardises the complexity increment of the model and computational effort without taking advantage of that in final accuracy. As mentioned, a particular type of MLP was used, and the difference consists on the connection between the hidden layer and output neurons, a well-known classifier, **Logistic Regression** made this connection.

As MLP serves to downsample results of CNN application, Logistic Regression was able to give probabilistic interpretation to an output of MLP and on it resides the classification task attributing a probability to a **unknown event**. One of the essential functions that able this probabilistic interpretation is *softmax* 2.15, giving value in a range of 0 to 1 to a particular event. In the topology classification problem addressed here, means a probability of two possible events, the biggest one will be chosen as the correct identification of status. Also, the method conserves the independence of events, and a breaker cannot be closed and open at the same time, meaning the sum of probabilities resulted from the classification is equal to 1.

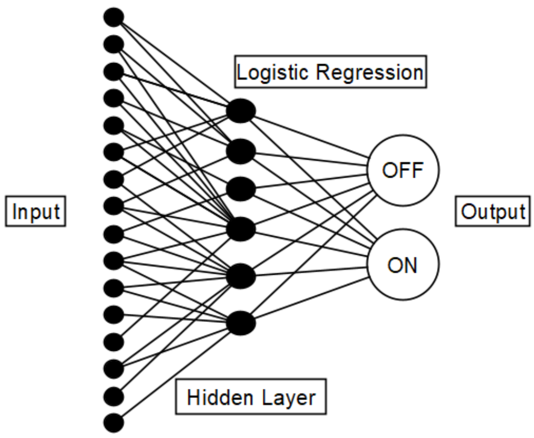

Figure 4.3: Proposal Classifier with demonstrative input values to achieve a binary classification.

Finally, the typical architecture of the used classifier in the developed work is presented in 4.3, and topic 4.2.3 will be dedicated to the supervised training and evaluation of solution in an iterative procedure as well as principal concerns of that.

4.2.2 CNN structures

After defining the classifier that will be exercised on CNN results, it is still necessary to search for ideal configuration of the layer's, including the number of them, filters size and pooling operators dimension. As some of Deep Learning frameworks, a rule model to determine such construction is not clear, and some tests were performed in order to produce an adequate model. The main learned idea was the importance of establishing a trade-off relationship between several input values and free parameters to be optimised as well as deeper structures that have an upper bounded level of layers regarding input matrix dimension.

As others CNN applications prove the benefits of using squared matrix, presented models will focus on that foundation besides as the amount of available data, almost the totality of variables (121 values of 124 possibles) define an 11x11 matrix and the possibility of a fewer quantity of measurements, 36, resulting in a 6x6 input dimension.

Table 4.1: Specification of developed models, each layer and variable parameters.

CNN layers	Proprieties	Model A	Model B
Input	size	11x11	6x6
1st convolutional	no. of kernels filter shapes	6 (3x3)	15 (2x2)
1st down-sampling	no. of kernels pool size	6 (1,1)	15 (1,1)
2nd convolutional	no. of kernels filter shapes	8 (3x3)	20 (2x2)
2nd down-sampling	no. of kernels pool size	8 (1x1)	20 (2x2)
3rd convolutional	no. of kernels filter shapes	8 (4x4)	-
3rd down-sampling	no. of kernels pool size	8 (2x2)	-
Fully-connected	input units hidden units activation function	32 6 ReLU	80 6 ReLU
Logistic Regression	input units output units activation function	*6* *2* *softmax*	*6* *2* *softmax*

Some tests were produced in order to attain the best model A and B presented on table 4.1 and applied to proposal inputs matrices, main consideration based on a trade-off between computational effort and accuracy model testing. Thus, a very complex modelling also produces a more challenging training procedure with significant run times, not being beneficial to practice applications.

Comparing each model, Model A allows a deeper structure with three convolution and pooling layers due to more input variables (121) contrasting to Model B that at most reaches 36 possible input measurements and so that only two layers can be attained. Also associated with the number of inputs, is the number of filters and the length of CNN section. Model A and B represent such differences. On performed tests the number of filters was increased until a stage of model maximum tested accuracy, that meaning if an increment of filters were made, the improvement of performance was not observed and sometimes it deteriorated.

However, also the down-sampling phase has some considerations in order to preserve a depth CNN model. Initial pooling operations only copy the result of convolution operation, preserving the number of filters and values and final layer of each model has the most responsible step of features recognition. Applying a travelling 2x2 matrix to previous results of a convolutional layer, it will choose the most significant value of 4 possibilities, then a feature is collected and is ready to serve as input to classifier stage.

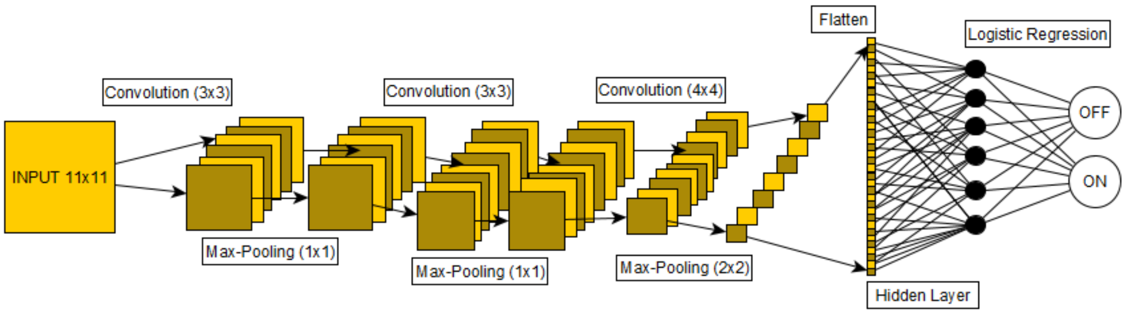

Figure 4.4: Layout of CNN structure to 3 layer example with principal operations used in the classification problem of breaker status recognition, ON or OFF.

Projection of depth CNN arrangements was made, considering the possibility of performing such models on challenging situations and consequently scenarios with poor information content about breaker status. Figure 4.4 demonstrates exactly the diagram of model A with the most important operations to facilitate comprehension of the main idea implicit to the developed topology processor. For model B, representation is similar to the model A changing the number of filters and layers of CNN stage. Later, an explanation of model training execution and performance evaluation will be introduced as an essential part of the obtained results.

4.2.3 Training Procedure

The neurons presented on MLP as another type of neural networks have incorporated a nonlinear activation function, hyperbolic tangent, *sigmoid* and recently ReLU as most similar to neuron biological behaviour, all of them will be tested. This type of neural network also is subjected to training trying to achieve adequate free parameters -weights and biases - and so that training procedure results on a paramount concern as well as the CNN modelling.

Projected models were subjected to a most used supervised learning technique, **backpropagation**, working as a feedback index on training procedure to adjust weights and biases. Essentially this algorithm divides into two main steps, **error propagation** and **free parameter adjustment**. Firstly, the matrix input feeds the model and go through model layers, reaching the final layer where classification result is obtained and compared with desirable output. Thus, a comparison between them is made by a **loss function** resulting in an error that in a second stage, act as an indicator to adjust all free parameters resided on layers.

4.2.3.1 Loss Function

Evaluation of accuracy on training execution as mentioned, it is in charge of a loss function or also known as cost function due to the impact that affects the training procedure. In the classification task performance, a most straightforward function is usually applied, **Zero-One cost function**, to identify erroneous classifications. On modern topology problem approach, that means a false open breaker status and contrarly an untrue close state.

Such function gives a qualitative point of view to the definition of switch mode. If it is correctly defined or not, but a model that produces an erroneous classification on a specific scenario could be significant, and being far away of correct status or close to the right answer. This problem induces on the necessity of a fine-tuned cost function that is capable of bringing feedback of the classification task. Thus, that emerges **Negative Log-Likelihood (NLL)** giving a quantification of how far is the model, in free parameters definition, to produce the correctness solution.

NLL loss function will be applied to results produced by *softmax* function 2.15, and that means the application of logarithm function to a range of values between 0 and 1. An expression that defines the NLL can be represented for a set of evaluated classes x with input probability y:

$$NLL(y,x) = -\sum_{i=0}^{x} log(y_i) \tag{4.1}$$

The interpretation given by NLL functions is not more than convert input probability to a specific group of values that provides confidence to the predicted class. In other words, a class with higher probability result of softmax operation with NLL analysis will make a low-cost value and contrarily, a lower probability produce a higher loss result, given a particular confidence degree. Developed models use NLL as loss function trying to attain the adequate free parameters and model optimised by the cost function results.

4.2.3.2 Mini-Batch Gradient Descendent

After the evaluation stage by NLL loss function, optimisation of weights and biases inherent to each layer is based on the gradient descendent procedure. This technique tries to achieve a global minimum optimal point of the cost function, defining an adequate operation point linked to the input available on an iterative training. Mathematically, the gradient descent optimisation algorithm to weights (W) and biases (b) can be described by:

$$\begin{cases} W_i = W_{i-1} - \alpha \cdot \nabla J_i(W,b) \\ b_i = b_{i-1} - \alpha \cdot \nabla J_i(W,b) \end{cases} \tag{4.2}$$

New parameters are obtained with a derivation of loss function resulting in gradient $\nabla J_i(W,b)$ where it means the direction that value x assumes. Then the convergence to a minimum local

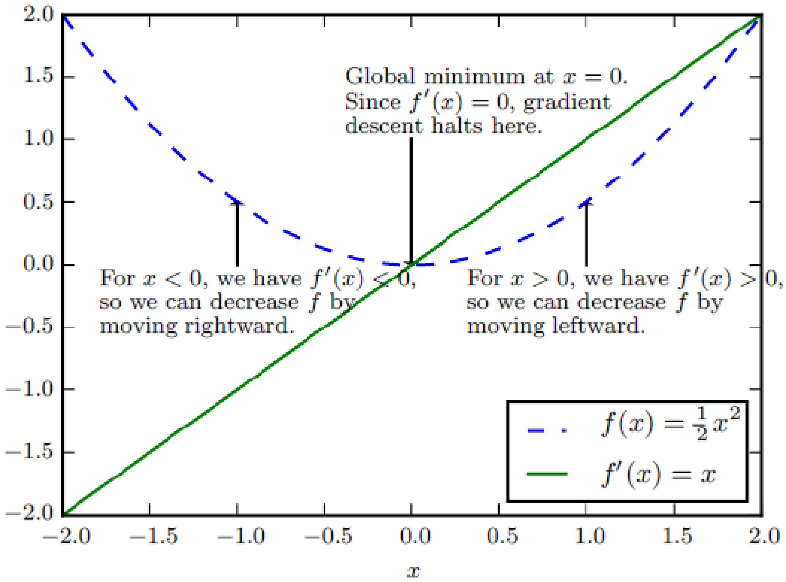

Figure 4.5: Illustration of the gradient descendent technique used to a single input function $f(x)$ with path demonstration attaining global minimum on an iterative procedure [39].

error means appointing to the opposite direction and so that $-\nabla J_i(W,b)$ ensures the catching for a minimum of the loss function, even if any position that x can occupy in the domain 4.5.

The transformation that affects the new parameters generation also is affected by a learning rate α to a smooth approach to a minimum error. Introduction of a learning rate avoids big jumps of weights/biases values as a result of the left and right curves of the loss function alternation 4.5. On the worst case standing in that alternation for a long time without search for the optimal point. The training procedure assures a learning rate $\alpha = 0{,}1$ to all presented tests.

Considering an extensive dataset, running an epoch of training execution and evaluate the impact of each input to the parameters optimisation takes much time and cause an effort task. Still, the mini-batch is studied as the adoption of a small number of samples and the average of each cost function result are considered to the gradient descent algorithm application. Thus, a reduction of the procedure run time is taken into account, and mainly, implementation of mini-batch samples reduce the probability of it been stuck on a local minimum and escaping to a desirable minimum loss function value. All the produced tests incorporate the batch-size of 30 samples, despite that fact, a more significant number of samples can be applied, decreasing the training time duration, although a most accurate model cannot be assured.

Finally, a diagram of an i epochs of training procedure is presented on figure 4.6 with value parameters, W and b, randomly initialise and subjected to n input samples on an iterative execution until loss function evaluation determines his minimum error.

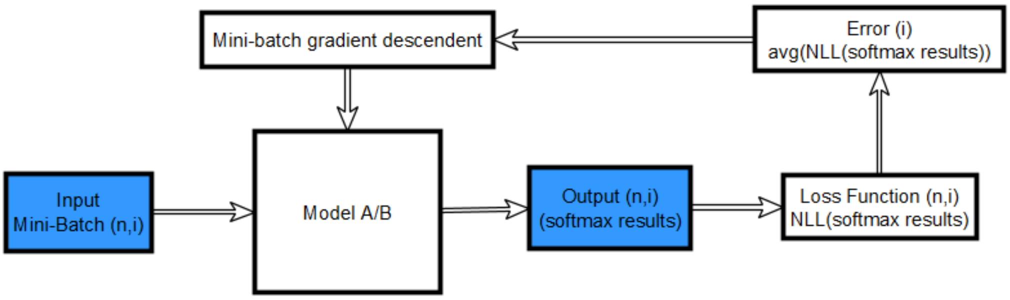

Figure 4.6: Diagram representing i epochs of iterative procedure with batch size n.

4.2.3.3 Training problems - Regularisation

Implementation of a training procedure on propose models face one of the most frequent problems the **overfitting** of trained data as well as the general Deep Learning frameworks. This issue occurs when the trained model produces a low error, yet, comparing to test set evaluation the accuracy differences are evident. On the overfitting phenomenon, the trained model produces an error lower than what is observed in the test set. Such a difference is so significant that is possible asserts that the trained model over adjusting to data samples 4.7.

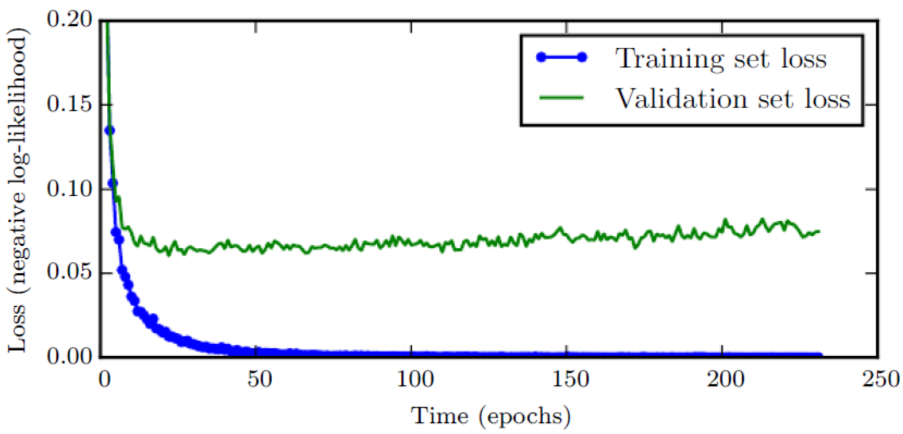

Figure 4.7: Iterative procedure representing overfitting event along with epochs number increases [39].

Along with the history of Deep Learning frameworks development, many strategies were adopted to escape from overfitting zones, and the present work enforces one common approach: **early stopping procedure**. Firstly, the early stopping emerges as the most straightforward technique in a training/validation/test algorithm based on the analysis of the training procedure. It means where a validation set is performed to corroborate the trained model and reveals one worse success rate than the previous epoch, the training procedure stops. Implemented strategy can act differently, stopping the iterative process immediately or adopt an adequate patience view where training model waits x executed epochs without observing training improvement. If such condition was verified, the run algorithm ends and assumes that it will not converge to better modelling. The mentioned approach reveals a proper idea allowing the progression of the iterative procedure without taking in cause non-desirable early stop. For the presented topology processor, although that technique seems too simple, it acts efficiently without interfering directly with the training execution, and to primary studies, this is the adequate regularisation technique.

4.3 Input structure organisation

A CNN structure requires a 2D input, and such characteristic was the main reason to choose CNN as a framework to the switcher status classification problem. The input image also can be considered as a matrix of numerical values instead of a 2D picture. Thus, the position that each variable will occupy on matrix influence the extraction of unseen features by the convolution and the pooling operations. Despite a random place of input measurements can be verified, yet that fact disables the ability to use the CNN greatest propriety - the vicinity correlation between each pixel (value) to facilitate extraction of features.

1	3	5	11	17	27	37	51	65	83	102
2	4	7	13	19	29	39	53	67	85	104
6	8	9	15	21	31	41	55	69	87	106
10	12	14	16	23	33	43	57	71	89	108
18	20	22	24	25	35	45	59	73	91	110
26	28	30	32	34	36	47	61	75	93	112
38	40	42	44	46	48	49	63	77	95	114
50	52	54	56	58	60	62	64	79	97	116
66	68	70	72	74	76	78	80	81	99	118
82	84	86	88	90	92	94	96	98	100	120
101	103	105	107	109	111	113	115	117	119	121

Figure 4.8: Input structure with 11x11 dimension and organised by D_{CS} criterion where 1^{st} represents measurement with greater distance (most content representation) until to 121^{th} position, value with the lowest distance and less contribution to breaker status definition.

Hence, the proposed structure 4.8 was made considering distance content measure explained in previous chapter 3, where each available variable represents a degree of importance to the breaker connectivity - ON or OFF – exploited by a 2D pattern definition. So, the main idea is organising most correlated measurements next to each other, as an attempt of establishing most recognisable patterns to CNN built models.

The generated dataset 4.1 defines 124 possible power flow results – lines power flow and power injections – however, it is not available the definition of a square matrix with 124 values. On the literature review, it was mentioned the advantage of the square matrix to attain a better-trained model. To achieve this goal, and trying to preserve the maximum number of variables, an 11x11 dimension matrix defines the usage of 121 measurements. The elimination of 3 input values is not problematic, and the global network is properly represented on the input matrix.

Figure 4.8 with heat map representation indicates the matrix values arrangement, where the most valuable was located in the $1st$ position (top left corner) and the content less informative values was organised in the vicinity of each other until the $121th$ measurement (down right corner) with less information about the switcher status. The chosen architecture to input values also provides an easily multidimensional definition preserving intact the order idea. In other words, for example, it was possible takes from 4.8 to defines the 6x6 input 4.9 just with 16 values. Such scheme rejects the remaining twenty values to produce some tests with lack of information and still using the CNN defined models.

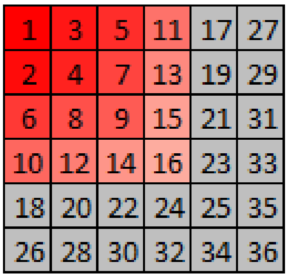

Figure 4.9: Input structure with 6x6 dimension and organised by D_{CS} criterion where 1^{st} represents measurement with greater distance (most content representation) until to 16^{th} position, value with the lowest distance and less contribution to breaker status definition comparing to 1^{st} value.

Past some executed tests to a reduced input, was denoted that to conserve the number of convolution and pooling layers of the model B and test less than 36 input values, it is necessary rejects some variables. The size of filters from CNN's layers on model B has a compact size, and it is not possible to rearrange them for an input matrix with dimensions less than 6x6 4.9.

Thus, a constant value out of input range measurements was added to all scenarios in tests that were used less than 36 values on model B. A introduced constant value in practice represents input measurements without signification to the breakers status recognition. This implemented idea does not take severe problems in obtained accuracy of the trained model – the pattern of values is always presented - and the CNN will be able to recognise them correctly.

4.4 Substation internal topology

Addressed reconfiguration problem of the single breaker status extended along the power network was the starting point to a more sophisticated approach - the determination of internal topology of a substation. This task can be seen as the most difficult on configuration arrangement problems in the power network. A substation represents a particular point of the system where a considerable amount of n breakers with 2^n possible configurations. The independence between each breaker incorporated on substation is what determines such amount of topology combinations.

Naturally that each substation operates in a specific arrangement, the most commonly used are single-bus, double-bus (or double-breaker) and breaker-and-a-half. However, it is necessary to consider the possibility of any configuration, with possible problem occurrence on a breaker and conduces to a not expected topology configuration. The identification of unexpected changes based on the information around the substation is an essential part of the developed work. Thus, it was considered two different substations as study cases:

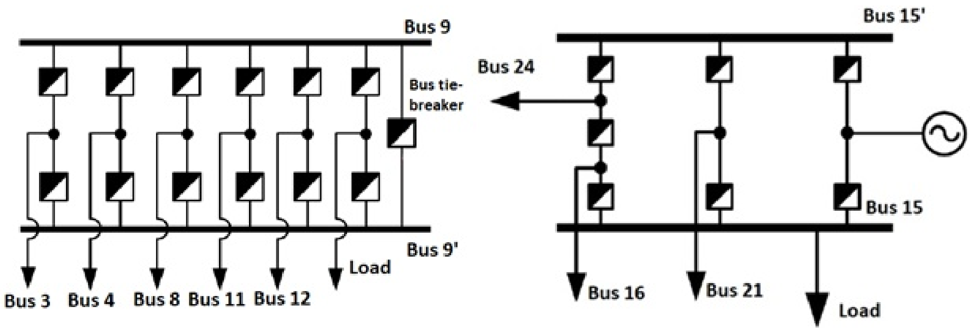

Figure 4.10: Scheme of internal topology breakers of the substation located on bus 9 (LEFT) and bus 15 (RIGHT) with connections to respective buses.

Chosen substations were based on the number of breakers that each one can incorporate. Thus, substation 9 with more possible topologies defines a more complex problem, and oppositely with just seven incorporated switchers, substation 15 characterise a more manageable problem.

Considering the classification problem showed on 4.2 to a single breaker status determination, the same such proposal will be applied to present substation topology recognition. An adequate CNN model will define each breaker to determine the connectivity of it on a particular substation point. Evaluation of information presented on topology variables inherent to each breaker status of the substation will also be performed with probabilistic distance D_{CS}. Then, as single breaker analysis, on substation problem, each breaker are linked to specific variables of the network defining a unique identity.

The generated dataset also considers the same load variation conditions and noise introduction demonstrated on 4.1, but in this specific case, a 25^{th} bus was integrated representing the unfolding of the bus where the substation is located. A representative amount of topologies was considered to the different substation cases determining a representative dataset of 20000 samples. Each possible input represents the switch mode - ON or OFF - to every one scenario with power flow measurements also as resulting power injections.

Although the n CNN models will perform the accuracy of the breakers, due to independence between it, and being $b_i(\%)$ the efficiency of each switcher, the global performance of substation identification is defined by:

$$Substation_{performance}(\%) = \prod_{i=1}^{n} b_i(\%) \tag{4.3}$$

Following chapter will show the results of CNN models application to these substations with some considerations, different experiences based on informative zones and how to achieve new degrees of information content to perform Deep Learning models as CNN.

4.5 Final Remarks

This chapter of the dissertation pretends to clarify the focused problems of this document. Also, the used techniques to determine CNN structure, the training procedure and input values alignment are described carefully to offer a clear vision to the reader of the advantages that suggested models can bring.

Additionally, it is essential to mention that generated datasets of the test system to single breaker reconfiguration, and substation topology problem was provided by Jakov Opara and used to apply developed algorithms. Subsection 4.1 emerges of an additional explanation necessity and awareness to the importance of data samples as well as the generation of real values with mentioned proprieties of load change and noise introduction.

Presented models were developed on Python programming language where construction, training and test of them was achieved. Directly associated with Deep Learning frameworks, the produce of showed CNN take into account three essential libraries: Numpy, TensorFlow and Theano.

Chapter 5

Results

After the methodology clarification on the previous chapter, now is time to validate the presented ideas starting for dataset definition and followed by CNN models definition, the relevance of input values arrangement on matrix spatial representation and principally the definition of CNN as a properly topology processor.

Such initial models validation is necessary to prove the following results: the test of 10 single breakers of a transmission grid, the substation topology configuration problem and meters optimal location. The CNN modelling was guided by an informative probabilistic contribution of distance D_{CS} defined in chapter 3. Also, the reader is advised to the comprehension of the chapters 3 and 4 to understand the showed results.

5.1 Initial considerations - Dataset

According to several works done on Deep Learning field, the definition of dataset splits is a crucial step to attain a correct trained model and most importantly, a test of it with real targets to evaluate these same models. The division of data on training, validation and test sets has not a defined rule. For a higher number of scenarios as the 20000 samples of the present problem, the conventional division of it is the following:

- **Training Set:** 70% of the total data are addressed to the supervised training, in order to attain the best model that reproduce the known output values;
- **Validation Set:** 15% defines the samples that are used to validate the free parameters, weights and biases of the iterative modelling;
- **Test Set:** The remaining 15% of the global dataset is reserved for testing the trained model and measure the accuracy of it on unobserved scenarios in order to define the efficiency of the model;

Such division of input cases is adopted on single breaker problem data division, as well to the substation topology issue too. Researchers with published works about Deep Learning frameworks also purpose other divisions, for example, 60% to trained data, 20% to validation set and 20% to test samples. Also, it mentioned that for a more significant dataset multiples ideas are well-founded, since the representativity of these samples on each group of data is present, principally on a test group of unobserved scenarios.

This same aspect was assured on first sets of experiences, where different test samples were sorted and tested, and the most representative group of cases was assumed to the next experiences of this chapter.

5.2 Single Breaker Analysis

The present section will be addressed to a single breaker classification problem, and initially, some experiences were done to define most adequate models to perform this same task. After this first exhaustive step, a group of tests will be clarified trying to guide the reader to understand some inherent aspects of topology estimation. The importance of the correlation between power flow measurements is shown as well as how CNN is the most accurate framework, with an impressive pattern recognition efficiency. Principally, these same aspects will be proved in a realistic operation scenario with a lack of observability.

5.2.1 Corroboration of models

A crucial part of developed topology processor was the focus on the study of CNN existent models, mainly how it works and how it is possible to adapt this framework to the classification of breakers localised on a generalised network. Some configurations can be changed to define a properly CNN architecture. These same proprieties are:

- Normalisation or standardisation of the input values;
- Model depth (number of layers);
- Size of convolutional filters;
- Number of features incorporated on a convolutional layer;
- Classifier size and structure;
- Hidden layer activation function;

Although the non-definition of a rule model to define a CNN structure, the study of some inherent aspects is necessary. Despite that, some proprieties as the size of convolutional filters, the model depth and the hidden layer size was defined with several test tries and did not represent a scientific definition behind it. For the present problem, it is not possible to assure that a deeper CNN structure will achieve better results than a less one, as well as the size of filters are not directly linked to the final efficiency of the classification task.

Still, with gained experience of performed tests, some crucial considerations were taken into account and were observed that some configurable proprieties of developed topology processor are directly associated with the global performance of training procedure (run-time and accuracy). On the present dissertation, it is not possible to show all the performed tests that guided to the definition of the best models. Although, it will be clarified three main ideas: the influence of free parameter number (weights and biases), the definition of input data treatment and the activation function of the hidden layer classifier. For that purpose, figure 5.1 will be shown the evolution of the training procedure of breaker 9 classification under model A 4.1 application. It represents the best modelling to define a point of reference to the next examples of possible changes that the architecture of the neural network can exhibit.

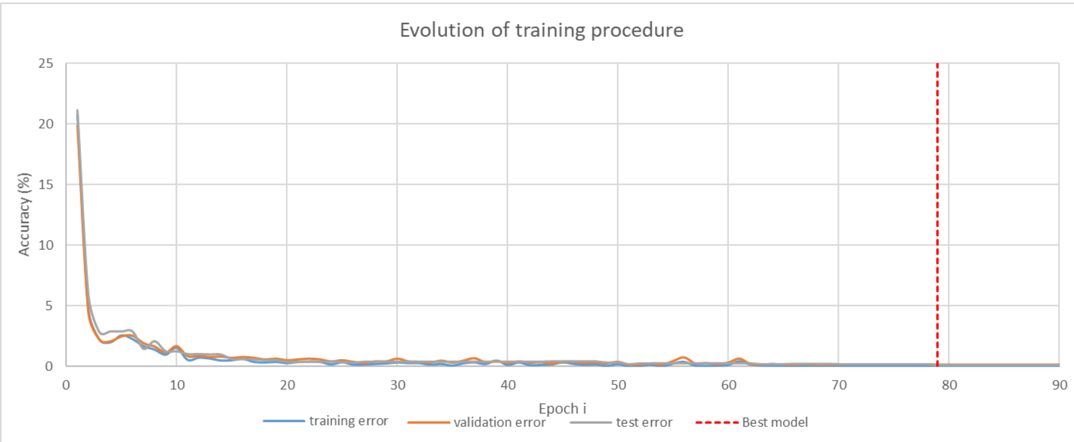

Figure 5.1: Error accuracy of training procedure for breaker 9 classification using model A and considering 121 available measurements as the non-organise input values.

As it is possible to observe on graphic 5.1, the training procedure for breaker 9 example performs a soft convergence to the best model on epoch 79 with inserted inputs perfect trainable. Also, it is notorious the non-overfitting of data samples with the validation set error staying equal to the test error after finding the best model. Such example demonstrates a correct trained situation where a satisfactory precision was attained. Although, with the increment of convolutional features number, it is possible to imagine that a better performance can arise. Such idea in practice does not occur, to prove it, one of the various tested models is defined in table 5.1 as the model A_2 comparison with model A (A_1).

Table 5.2 displays the efficiency of these two models, and a better precision was achieved on the first one with fewer features. Despite the neural network A_2 converging to the best model with fewer iterations of the training procedure, the number of parameters to be optimised is enormous comparing to model A_1. On this example, an extra computational effort represents a time execution 7 times higher than model A_1. Accuracy differences between them are not significant, although, the vital aspect to chose model A_1 resides on the computational effort concern, it reveals to be the best option for all 10 tested breakers.

Table 5.1: Architecture of neural networks model with a different number of free parameters integrated on it.

CNN layers	Proprieties	Model A_1	Model A_2
Input	size	11x11	11x11
1st convolutional	no. of kernels filter shapes	50 (3x3)	6 (3x3)
1st down-sampling	no. of kernels filter shapes	50 (1x1)	6 (1x1)
2nd convolutional	no. of kernels filter shapes	20 (3x3)	8 (3x3)
2nd down-sampling	no. of kernels filter shapes	20 (1x1)	8 (1x1)
3rd convolutional	no. of kernels filter shapes	20 (4x4)	8 (4x4)
3rd down-sampling	no. of kernels filter shapes	20 (2x2)	8 (2x2)
Fully-connected	input units hidden units activation function	80 10 ReLU	32 6 ReLU
Logistic Regression	input units output units activation function	6 2 softmax	6 2 softmax

Table 5.2: Comparative results between models with a different number of free parameters integration.

Breaker 9	Model A_1	Model A_2
no. of failed tests	4	5
Epoch of best model	79	15
Failed tests (%)	0.133%	0.166%
Accuracy	99.87%	99.83%

Another crucial point to take into account is the definition of neural activation function of the hidden layer that connects the CNN pattern extraction to binary output (ON or OFF). The table 5.3 easily expose the efficiency of different function options where ReLU performs the most accurate classification.

Table 5.3: Performance of model A on breaker 9 classification problem with three different neural activation functions: ReLU, hyperbolic tangent and sigmoid.

Activation function	ReLU	tanh	sigmoid
no. of failed tests	4	11	14
Epoch of best model	79	70	59
Failed tests (%)	0.133%	0.367%	0.467%
Accuracy	99.87%	99.63%	99.53%

Figure 5.2: Error accuracy of training procedure for breaker 9 classification using model A and consider 121 available measurements as the non-organise input values.

Focusing on the executed test where a sigmoid function was used on hidden layer neurons 5.2, and it is observed the efficiency irregularity of models on epoch *i* with considerable changes on the successive iterations. This same procedure evolution nature is present where the hyperbolic tangent was used as an activation function. For these two functions, it was revealed the incapacity of these to conduct the CNN output values to fed Logistic Regression operation comparing to ReLU function. Also, to a procedure with more than 100 iterations, the training evolution goes to an overfitting area with validation error growing to higher values. The described event corroborate the recent researches that prove ReLU as the most similar modelling of neuron behaviour on processing information between them.

Finally, another essential aspect mentioned and is not directly inherent to CNN modelling is the preprocessing of data before it serves as the neural network input. On different fields of data science, the data treatment takes immense importance. The Deep Learning area background usually defines the normalisation and standardisation of inputs data as a proper way to feed the neural networks. Such an idea addressed to present problem was tested, and graphic 5.1 defines the execution of model A to standardised inputs. The input data normalisation also was tested, and on next illustration is possible to see the breaker 9 classification procedure using model A with normalised inputs on figure 5.3.

Differences between these two data methodology treatment are notorious, under normalisation inputs was observed that to most of the breakers, the training procedure cease on first iterations without progression of it. Such incapacity of growing to a most accurate model defines normalisation of data as the worst type of data representation comparing to standardisation.

Over the present section, it was presented important aspects to modelling a CNN correctly with the explanation of the model A definition. For model B, the same ideas were taken into account. The principal difference associated with it is the less number of input values, and as well minus one convolutional layer that allows the definition of more pattern features without redundancy associated with it. Thus, to this second model, the idea of a trade-off between model efficiency and a less time execution guides that construction.

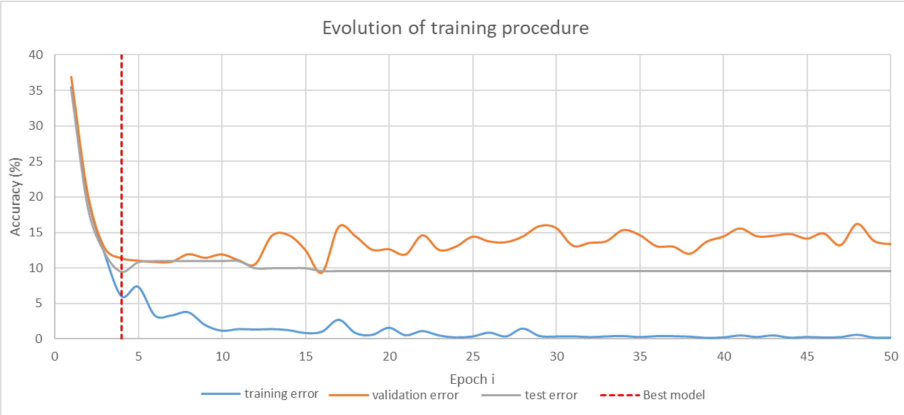

Figure 5.3: Error accuracy of training procedure for breaker 9 classification using model A and consider 121 available measurements as the non-organise input values. The input normalisation was made on a range of $[-1, 1]$.

5.2.2 Influence of input arraignment

After the definition of Models A and B, the present subsection introduces firsts tests applied to a single remote breaker on a network. For all 10 switchers, to prove the ability of presented CNN structures, it was used Model A considering all available measurements and without a specific 2D matrix arrangement. The present test is essential to define if such Deep Learning framework can handle with this topology classification problem. On the Data Science field, each problem has a cluster of tools that could model it, and here, a unique proposal method is presented requiring such clear demonstration.

Table 5.4: Reconstruction of the 10 breaker status with 121 values input using Model A and also with an equal non-defined organisation measurement.

Breaker	1	2	3	4	5	6	7	8	9	10
no. of failed tests	0	0	6	0	1	0	1	10	4	0
Epoch of best model	5	6	48	8	2	1	2	87	79	3
Failed tests (%)	0.0000%	0.0000%	0.2000%	0.0000%	0.0333%	0.0000%	0.0333%	0.3333%	0.1333%	0.0000%
Accuracy	100.00%	100.00%	99.80%	100.00%	99.97%	100.00%	99.97%	99.67%	99.87%	100.00%

However, in the first experience, the potential informative content that measurement could express is not taken into account. Despite that, to each breaker was removed 3 most distance measurements from breaker localisation. That elimination of input variables is linked to the necessity of using 121 (of 124 total) input values to fit on the Model A input matrix.

The results on 5.4 show an impressive reconstruction capacity of the switch mode by Model A with a perfect definition of breakers 1, 2, 4, 6 and 10. Differently, the rest of breakers presents a total of 22 failed test cases with a higher impact on breaker 3, 8 and 9, taking to a longer iterative training. Although, the existence of scenarios with an erroneous classification, the global results are expected to define CNN as a structure that can extract good patterns on a matrix representation.

However, the non-definition of an input measurements arrangement represents the waste of all proprieties that makes CNN a particular framework on pattern recognition. Hence, inspired by the suggested input organisation presented on section 4.3 the second group of tests was performed based on that scheme. Accuracy of classification task applied under the mentioned conditions is present on the next table:

Table 5.5: Reconstruction of 10 breaker status with 121 values input using Model A and the matrix values organisation mentioned in section 4.3 of chapter 4.

Breaker	1	2	3	4	5	6	7	8	9	10
no. of failed tests	0	0	2	0	1	0	0	2	0	0
Epoch of best model	2	4	44	2	4	1	6	46	11	3
Failed tests %	0.0000%	0.0000%	0.0667%	0.0000%	0.0333%	0.0000%	0.0000%	0.0667%	0.0000%	0.0000%
Accuracy	100.00%	100.00%	99.93%	100.00%	99.97%	100.00%	100.00%	99.93%	100.00%	100.00%

Breaker status determination for previous experience proves the advantages of considering an arrangement of input data. The totality of failed identification scenarios reduced to 5 misclassifications, and it is expressed on breakers 3, 5 and 8. The introduction of that innovation presents to CNN model a more easily pattern definition, with particular attention to switcher 9, where the 4 missing classifications of the first test are now classified in fewer iterations. Such slight modification introduces an essential gain on method accuracy and also on training procedure, with the reduction of iterations number and consequently run time duration.

As defined in chapter 3, the probabilistic distance D_{CS} was used to infer the degree of information that a variable have about the status of the breaker. Figures 5.4 and 5.5 defines the top 16 of content relevance measurements, and two different situations also can be seen.

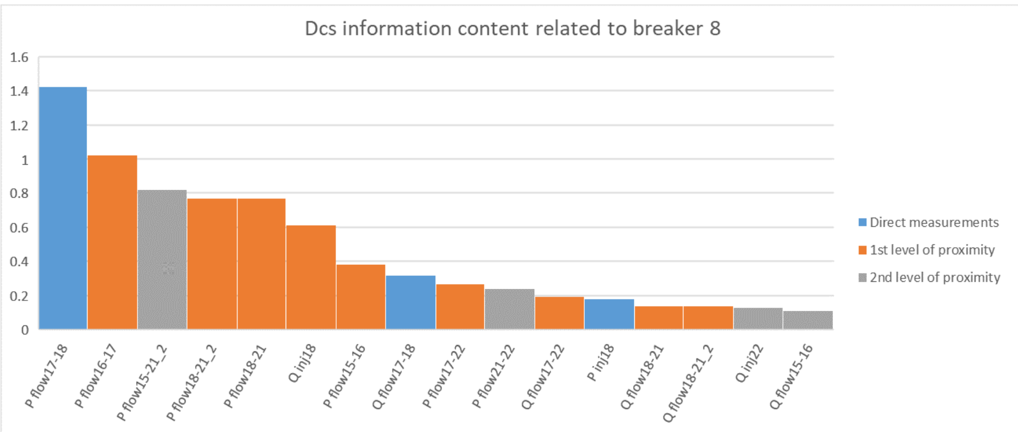

Figure 5.4: Breaker 8 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation.

Associated to breaker 9 is less significant information content on figure 5.5 than to the breaker 8 on the figure 5.4. First one resides on direct measurements and the parallel line power flow as the essential variables. The remaining variables and their pdf of open and close state define a non-distinctive relationship between them, so to the mentioned breaker, same measurements do not infer on their connectivity determination.

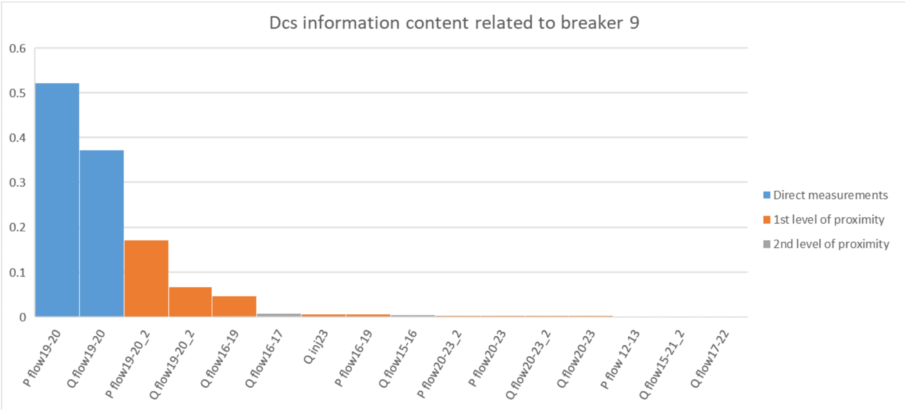

Figure 5.5: Breaker 9 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation.

Differently, to the breaker 8 probabilistic definition expresses more information about measurements, and it is possible to observe that direct measurements are not necessarily the most important. Even as Jakov Opara on their Maximum Mutual Information theory [35], here it is possible to analyse the power flow distribution across power network, sometimes engineering judgement as a decision criterion fails. As it detected on 5.4, from 4 possible direct measurements just active power metering on line 17-18, where breaker 8 is located, appears on the firstly ranked variables.

After proving the importance of input values organisation by the accuracy results 5.5, a visual analysis expressed on 5.6 can be demonstrated to bring confidence about the ITL techniques. The scheme of figures 5.6 focus on the 16 most essential measurements associated with the breaker 9 connectivity - left corner of 11x11 matrix. On the left column, the three different scenarios represent an open status, contrarily, on the right column is demonstrates closed scenarios.

The differences between - 5.6a, 5.6c - and - 5.6b, 5.6d - are notorious. Otherwise, there are similarities between them and for this 4 different scenarios of operation, it is visual a pattern that defines an open breaker and a closed one. Such evidence proves that the ITL definition could transform a group of signals on an image with proper identity - breaker connectivity characteristic - that a human being could visually detect such differences.

However, to similar operation scenarios, the difference between patterns that infers closed and open status may be not visually detected. To breaker 9, for example, it operates a specific line that is parallel to another one. So, if one of the lines are not connected, the energy goes through the other line without affects significantly the surrounding power flows. Thus, the input 5.6e and 5.6f represents that concern, focusing on the proposed organisation 4.9 and ranking 5.5, it is possible to see the main pattern distinction referent to the first position - the $P_{flow,19-20}$. Despite these similarities of operation scenarios, the CNN model can learn these patterns and defines a correct classification 5.5.

1.298998	1.296077	-0.21349	0.848628
-0.36452	-0.36322	0.737904	-0.41751
0.266562	3.822178	-0.60634	0.276498
0.846308	-0.41762	-2.06032	0.23567

(a) Open Breaker 9

0.000892	-0.15706	-0.30729	-0.47777
0.005853	-0.54622	0.402474	-0.32141
0.642431	1.076615	0.438072	-0.03855
-0.47901	-0.32539	-1.87972	0.860094

(b) Closed Breaker 9

1.826403	1.829973	-0.1205	1.425654
-0.38173	-0.37696	1.126094	-0.48979
0.711573	4.679441	-0.068	0.276311
1.417048	-0.49159	-2.58952	0.556147

(c) Open Breaker 9

-0.00088	-0.20379	-0.33578	-0.52477
-0.00144	-0.55007	0.428967	-0.32203
0.66267	1.116057	0.443739	-0.00406
-0.52954	-0.32437	-2.00962	0.729954

(d) Closed Breaker 9

0.561878	0.568779	-0.08641	0.266195
-0.15832	-0.15924	0.075795	-0.14898
0.294346	2.061745	-0.22137	0.587792
0.270587	-0.14804	-3.52501	0.275132

(e) Open Breaker 9

-0.00537	1.696975	-0.00497	0.476392
0.000182	-0.30446	0.22322	-0.22388
0.479399	2.702859	0.016542	0.397697
0.47578	-0.22436	-2.02503	0.148028

(f) Closed Breaker 9

Figure 5.6: The visual representation of the 16 most valuable measurements to define breaker 9 connectivity. Values were surrounding 0 p.u represents red tonality and module values bigger than it, is linked to yellow tonality.

Despite the performance on reconstruction scenarios of CNN model 5.5, it is essential to analyse the 5 wrong classifications trying to identify possible reasons to that happen. For that purpose, the failed scenarios for breaker 8 can be seen on 5.7 with active and reactive power over the line 17-18 assigned to closed situations as well as the false closed cases occurrences.

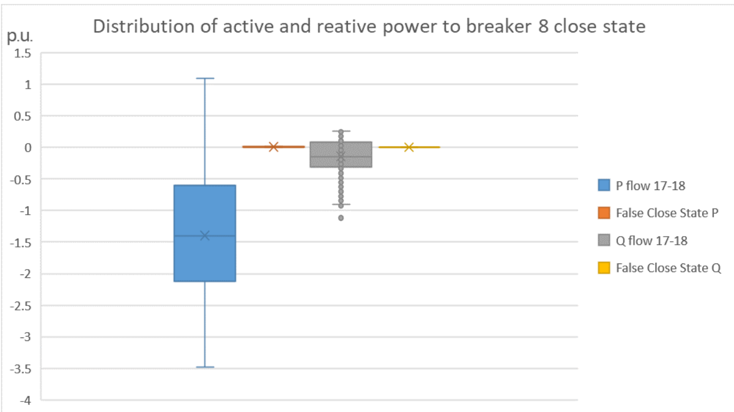

Figure 5.7: Representation of values that power flow on line 17-18 can exhibit and comparison with power flow of false close identification scenarios.

One of the most empiric acknowledgements correlated with switching mode is when power flow is higher than 0 p.u., and if the power flow measurement is correct, the line is connected with a closed breaker. Such a basic idea is right as well as it is possible defines when active power flow is near 0 p.u. the switcher is open, but it is not necessarily true. Lower values of active and reactive power flow can exist on a closed line, and that fact creates an ambiguous learning area to the CNN model. Also, other tested frameworks prove it [38], making the training of neural networks a problematic task when such events occur.

Both 2 failed tests identification on breaker 8 represents a false close classification, that means for each scenario input values are linked to an open switcher, and CNN classifies them as a closed breaker. Going to a more in-depth analysis of this case, it is possible to see the active and reactive power values near 0 p.u.. Observing graphic 5.6 where closed test scenarios are presented, similar situations to failed classifications are present and the ambiguous zone is defined here. The importance of $P_{flow,17-18}$ demonstrated by the probabilistic distance D_{CS} 5.4 determines this uncertainty, with the influence of that measurement being more significant than the remaining variables with the training procedure absorbing that characteristic.

Searching for similar cases than 2 failed classifications on the test set, choosing the worst failed case and so with a higher magnitude of power flow, it is possible to detect 24 cases similar to that on tested data. Therefore, the model A executes a proper classification of 22 scenarios with power flow near zero on line 17-18. The rest of missing classifications, associated with breaker 3 and 5, also are associated with a demonstrated example with lower local values. Also, on the switchers with 100% of test accuracy, the active power flow on it has significant relevance and scenarios with values surrounding 0 p.u. also appears with correct classification by the CNN model.

5.2.3 Performance under lack of measurements

Naturally that last performed tests do not represent a realistic and attractive application of neural networks, as mentioned, it was used to evaluate the capacity of CNN as topology processor besides the importance of considering a proper input organisation. In realistic scenarios of operation, the totality of information is not available. Sometimes just a few line meters are presented on transmission lines, the installation cost of it represents one of the reasons to a weak observable system, but another one is the failed telemetry or the possibility of meter damage.

Some tests were produced with fewer power flow results with the successive reduction of the input matrix, trying to achieve the less possible number of inputs and at the same time attain an excellent performance of models. Besides that, consideration of measurements linked to the breaker is not expected on a real application, the topology processor as a function of systems operators gain real interest when is necessary knows the connection of an untraceable line.

Inspired by this practical operation concerns, it was tested for all breakers without respectively direct measurements, power flow on lines and also power injection on delimited buses, and until 16^{th} most relevant measure associated to that breaker. Figure 5.8 displays the organisation of input matrices that feds CNN model B on training procedure without direct measurements, 1_{st}, 8_{th} and 12_{th} values of informative ranking 5.4 linked to breaker 8.

The application of the described idea was made using model B because it was projected to less than 36 input values. The accuracy of tests is identical using any of proposal models, and execution time are taking into account as the chosen criterion. Nevertheless is essential refer two special cases representing parallel lines (breaker 7 and 9) and to that cases, also is ignored the power flow of the adjacent line to represent a most realistic operation scenario. Test results can be seen on the table 5.6.

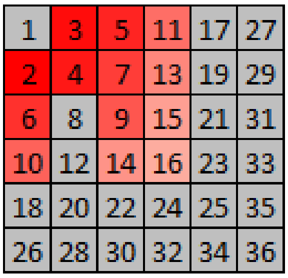

Figure 5.8: Illustrative scheme of input matrix organisation by D_{CS} distance criterion that feds CNN model B and determines breaker 8 status where coloured numbers represent available values and grey tonality the unavailable measurements.

Table 5.6: The accuracy results of executed tests under reduction to 16 possibles input values and without direct measurements.

Breaker	1	2	3	4	5	6	7	8	9	10
no. of failed tests	128	4	0	0	0	0	204	3	10	0
Epoch of best model	192	90	54	4	2	1	6	169	102	15
Failed tests %	4.267%	0.1333%	0.0000%	0.0000%	0.0000%	0.0000%	6.800%	0.1000%	0.3333%	0.0000%
Accuracy	95.73%	99.87%	100.00%	100.00%	100.00%	100.00%	93.20%	99.90%	99.67%	100.00%

Produced tests show a satisfactory classification accuracy, in some breakers remains 100% efficient, and the worst case is present on the switcher 7, but at the same time stay with satisfactory performance of 93,20%. As expected, the evaluation test set of some breakers in the fault of a considerable amount of measurements deteriorate his performance, and at the same time, the training procedure gets harder with the increase of epochs number.

A critical analysis of results 5.6 demonstrates a significant test error to breaker 1 and 7 compared with the rest of it. Thus, an electrical point of view is required and analysing the location of breakers on proposed test system 4.2, and breaker 1 is delimited for two PV buses with a load variation on each one. Breaker 7 is bounded by also two PV buses, and one of that, bus 15, with a significant load value of 317 MW. Hence, such lines are subjected to a changed power flow behaviour in a daily time operation, load changes impose the variation of generated energy on PV buses, and combination of this two aspects on the same bus makes this point of the network the force of power flow directions. Just the existence of power injection influence the power flow circulation, adding a variable load makes that highly unstable with a faster load change. In fault of direct measurements, these two breakers lost significant information about power flow of the network and probabilistic distance criterion D_{CS} shows the correlation between direct measurements to the status of breaker on 5.9 and 5.10.

Histograms 5.9 and 5.10 emphasises the breaker status dependency linked with direct measurements, being the most significance to topology classification procedure and, naturally, without these variables, the accuracy of models are compromised.

The previous group of experiences over topology processor model B was performed to a variable number of inputs according to each breaker, depending on if the best 16 variables include direct measurements on it. As illustrated on 5.8 with 13 input variables to breaker 8 trying to

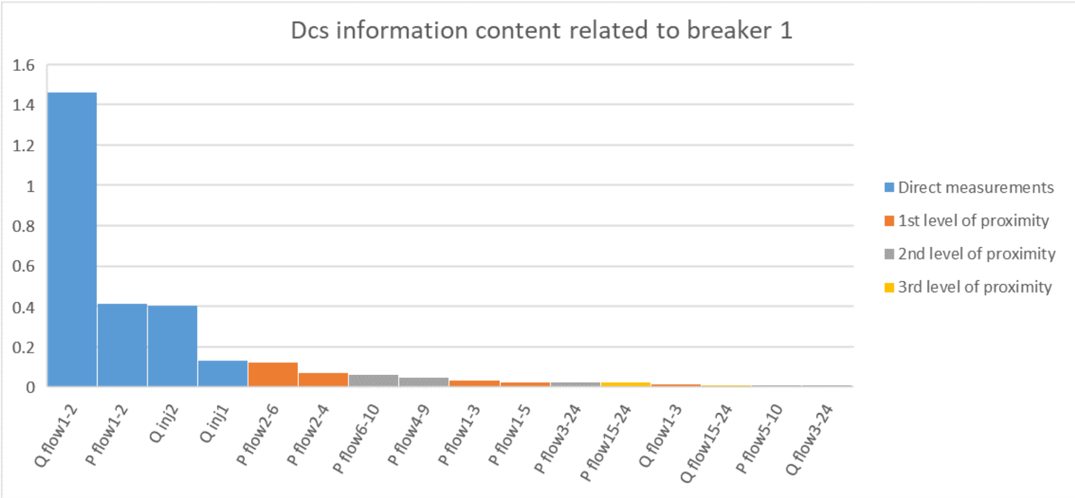

Figure 5.9: Breaker 1 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation.

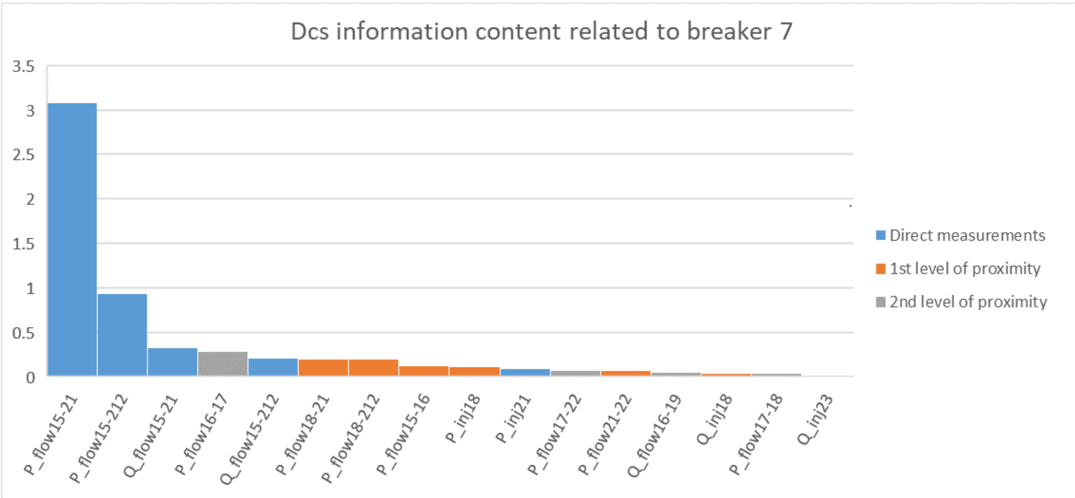

Figure 5.10: Breaker 7 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st} and 2^{nd} proximity levels from breaker localisation.

preserve spacial informative proprieties even if direct measurements are not present. Another idea was performed to switchers that maximum accuracy was not observed on past tests 5.6. It resides on without direct measurements, remakes the evaluation ranking to the new top 16 measurements adding successively earlier rejected values. The results of this approach are presented in the next table:

Such concept applied to breakers with the lowest information content as the breaker 7 and 1, including more measurements of network it is possible the improvement of success rate to levels upper than 95%. That change establishes a satisfactory result with a better pattern definition

Table 5.7: Accuracy results of executed tests under reduction to 16 most informative input values and without direct measurements.

Breaker	**1**	**2**	**7**	**8**	**9**
no. of failed tests	112	5	132	0	8
Epoch of best model	126	308	22	147	107
Failed tests %	3.733%	0.1667%	4.400%	0.0000%	0.2667%
Accuracy	96.27%	99.83%	95.60%	100.00%	99.73%

proved by reduction of epochs number to attain the best model on training execution.

However, the addition of more variables does not mean the achievement of a better CNN model, to breaker 2 for experience figured on 5.6 and 5.7 proves it. Incorporation of more power flow results on the input matrix also not adds significance to that breaker classification with worst pattern presentation to model B where is necessary 308 iterations to reach the best model. That evidence also is present when performed tests to input size reduction from 121 measurements to 16 5.6 on breaker 3 and 5. Firstly, where the totality of measurements was available, executed tests define wrong identification scenarios on these cases, but observing results of test 5.6 the perfect efficient is accomplished. Perhaps established idea of a more considerable amount of information contributes to robust models is refused here, and at this point redundancy of the number of variables was observed. Most of the times, reduction of input is a better definition to improves model optimisation without lost of efficiency.

That group of experiences proves to the breakers that are strongly linked to an essential group of power flow results that modelling topology processor can achieve extraordinary performances. Principally, even if only less than 16 measurements are available, and most valuable, without direct measurements. However, to breakers that connectivity status is not significantly related to measurements beyond the 1^{st} proximity level - without most important information - the classification task was involved, but an adequate efficiency can be ensured.

5.2.4 Realistic point of view

Until the present section of this dissertation to single breaker topology determination, some relevant tests were performed to understand the ability of developed models on achieving connectivity between two adjacent buses. Scenarios of lack of information were tested as well as the omission of direct measurements linked to the specific breaker. The local relevant information was taken into account with auxiliary probabilistic techniques, but a more realistic operation scenario is necessary, trying to evaluate these theoretic definitions on practical applications.

Thus, this section purpose the definition of a possible realistic network with 8 available local meters represented in figure 5.11 and with the location of the same breaker of previously performed tests.

Existence of 8 meters brings the possibility of considers 16 line power flows to feed the model B, and the same ideas of performed tests were admitted to this experience. Evaluation of informative content by a probabilistic distance D_{CS} was made according to each breaker, and input

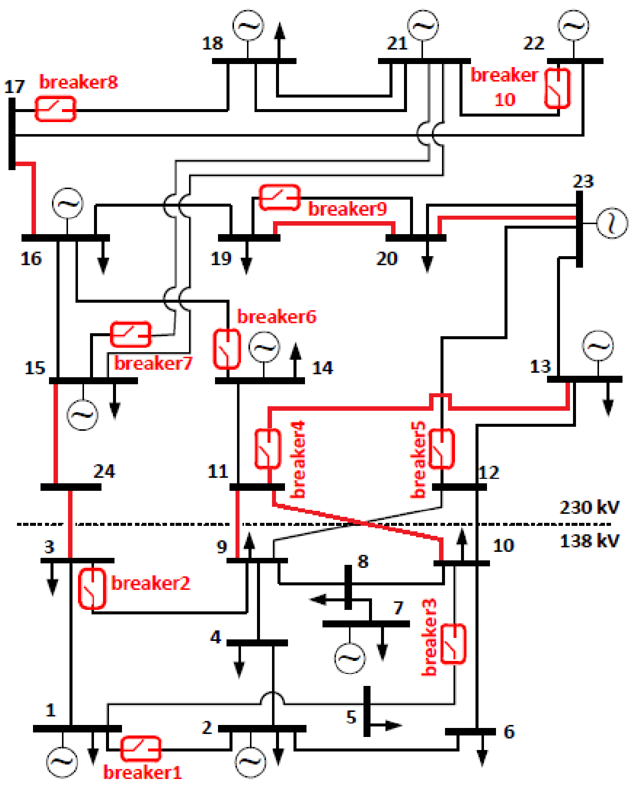

Figure 5.11: Breakers arrangement in test IEEE RTS 24-bus system where red lines represent the location of meters. The rest of lines are considered unavailable to performed tests.

organisation follow the same rule model defined earlier. So, it considers the order of the 16 measurements by a weighted correlation to a status breaker.

Described probabilistic evaluation tool brings a stronger acknowledgement to this introduced problem. Meters in some cases are not located beside the breaker area, and a lack of information about the network is evident. The also known engineering judgement to define most correlated variables to breaker status are not visible, and at most of the times, it is an impossible task over this traditional criterion.

Diagrams 5.12 and 5.13 express two most different scenarios, with a considerable amount of variables correlated to breaker 6 status than to breaker 2. The low distances D_{CS} defines a weak connection between the available power flow results and breaker 2. It is essential to observe that in these mentioned cases, direct measurements are not presented as well as other tested switchers. Except for one example, breaker 4 where local measurements are available also representing a realistic operation scenario with direct measurements but with lack of surrounding information.

Even with a not observable system, the ranking of measurements exploits some of power flow circulation trend based on analyses of historical background establish a correlation between different lines. On 5.13 it is possible to observe, as mention before, the impact of generation as the most critical parameter to defines the open and close status. So that, PV bus 23 from the

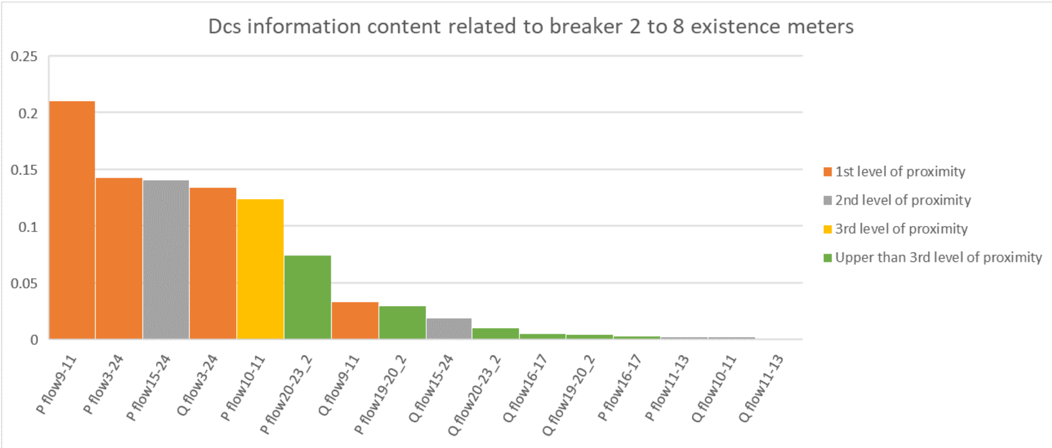

Figure 5.12: Breaker 2 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation.

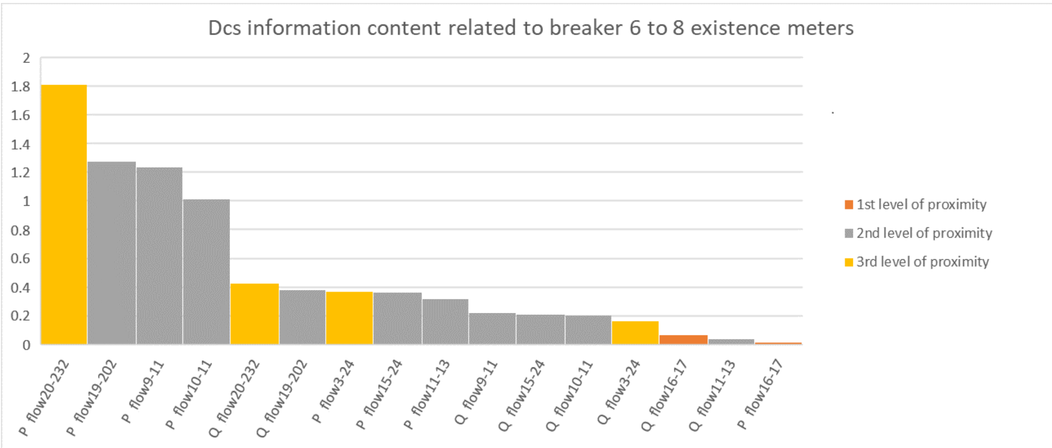

Figure 5.13: Breaker 6 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation.

available measure, are truly linked to the connectivity of breaker 6. First two most important measurements are $P_{flow,20-23-2}$ and $P_{flow,19-20-2}$ delineating the power flow path until line 16-14 where that breaker are located. The same behaviour could be found to breaker 2 analysis 5.12, even with inadequate available information, resides on measurements, the ranking of it shows the same importance of PV buses. On that case, 1_{st} and 3_{rd} most valuable power flow results of ranking 5.12 define the path until PV 15 bus showing a pattern of power flow through those lines.

Such a tool developed at this point of work shows how much powerful that could be to define unseen patterns figured on ranked measurements. Thus, CNN model B was the neural network scheme used to test such a realistic operation scenario. The accuracy of the results could be found in the next table:

Table 5.8: The accuracy results of executed tests under available 8 meters scenario to 10 single breakers.

Breaker	1	2	3	4	5	6	7	8	9	10
no. of failed tests	748	10	86	0	31	0	234	48	13	23
Epoch of best model	263	32	170	3	416	1	292	446	324	400
Failed tests %	24.93%	0.3333%	2.867%	0.0000%	1.033%	0.0000%	7.800%	1.600%	0.4333%	0.7667%
Accuracy	75.07%	99.67%	97.13%	100.0%	98.97%	100.0%	92.20%	98.40%	99.57%	99.23%

The performed test reveals an exceptional efficiency on a situation of observability system reduction. The existence of just 8 meters defines an acknowledgement of the $21,05\%$ about the totality of presented transmission network. As expected, CNN model applied on this test shows some obstacles related to training procedure with difficulties on recognising patterns under a complicated input nature.

Classification of the breaker status denotes an accuracy higher than 90% in most of the cases even as closest to 100% in a large group of switchers with breaker 4 and 6 reconstructions attaining such ideal goal. In different circumstances, the efficiency of breaker 1 connectivity classification was established on $75,07\%$ defining the worst tested case. Such breaker also in previously executed tests without direct measurements and under fewer measurements denotes some difficulties to produce an accurate classification.

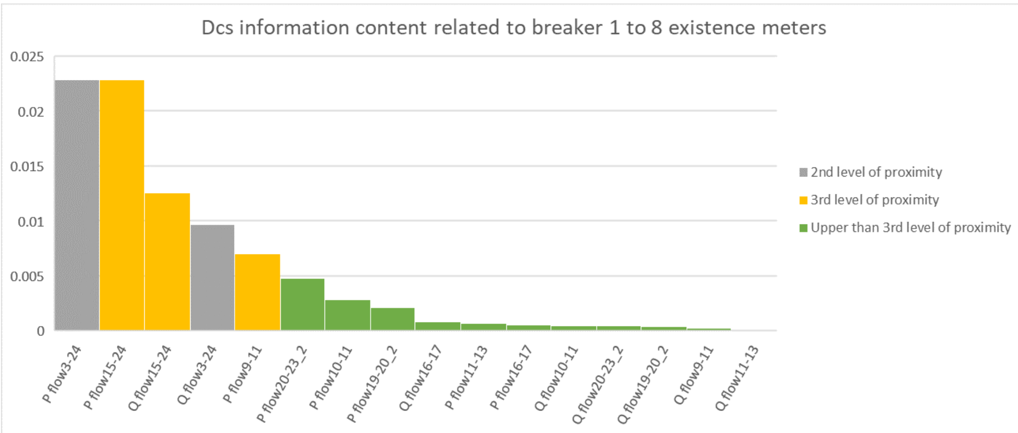

Figure 5.14: Breaker 1 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from breaker localisation.

Thus, going more in-depth on that case, it is possible to observe that breaker 1 location is far away of meters influence area 5.11. Naturally, that produced energy on buses 1 and 2 will supply near loads, as that is localised on bus 1, 2, 3, 4, 5 and 6, just on fewer cases, the mentioned generators will be selected to feed far away loads on the metering influence area. Before performed tests, with probabilistic pdf distance to arrange input matrices also was detected that for this particular case, the classification of breaker status would suffer from a significant lack of information as it is possible to observe on diagram 5.14.

Ranked power flow results correlated to breaker 1 defines as the essential meters which were located near to that point, in which power flow of lines 3-24 and 15-24 appear on firsts places of this ranking. Although the order of available results preserves the localisation to breaker 1, the probabilistic distance is lower, and that fact proves the incapacity of it to identify a possible connectivity state of the switcher.

Hence, proving that explanation, another test was done trying to understand if a closer meter can produce significant changes on the accuracy of this breaker. Naturally that the addition of direct measurements will improve classification of it, but this is not what is expected on real practice. So, introducing a meter on line 2-6 was tested removing the meter located on line 16-17 - a less significant one. Accuracy result goes up to 88,37%, an improvement of 13,30% comparing to the previous test with the introduction of measurements of the 1^{st} proximity level referent to that additional meter.

Increasing the supervised area of metering contributes to a significant improvement of the topology processor. On most of the times, a considerable amount of meters does not guarantee the efficiency of topology processors based on neural networks. A smart localisation of it can be achieved with a previous study of network background, and such task was demonstrated on the present dissertation. Also is essential refer that experiences of this section could not be done one classical topology processors, lack of observability system overlap the ability of such models to classify breakers connectivity.

5.3 Substation Topology

The first part of the work was to develop a group of CNN models addressed to the single breaker state determination with different scenarios tests understanding what is possible achieves with such ideology. Nevertheless, another topology problem emerges on grid operation, the substations located on the various buses incorporate a group of switchers that configure: the topology of the network, connectivity of loads and generators, if it exists on substation bus.

Substations topology is a complex arrangement of breakers, and it depends on the possible lines number that can reconfigure, with the number of switchers linked to that complexity. Although, those breakers operate independently between them, and that fact allows the interpretation of substation topology determination as a classification of a breakers group located on the same point of the network. Therefore, its connectivity is dependent on the power flow around the substation. Existence of power flow on a line is truly connected to those breakers ON or OFF status.

Inspired by the exhibited single breaker classification results, CNN model B was considered as the topology processor to perform tests on the substation problem. First was chosen a simple topology substation as it figured on 5.15 and after that, a complex substation with 5 possible connected lines 5.18. For both study cases, the dataset was supplied by Jakov Opara and generated following the same conditions to single breaker dataset creation. However, to this specific case, different substation topologies were considered trying to replicate potential operations scenarios of single-bus, double-bus as well as breaker-and-a-half configuration. That same concern defines 35 possible breakers arrangements to substation 15 and 211 topologies to substation localised on bus 9, establishing a global data set of 20000 likely operation scenarios, with chosen test group integrating all topologies in different network load variations.

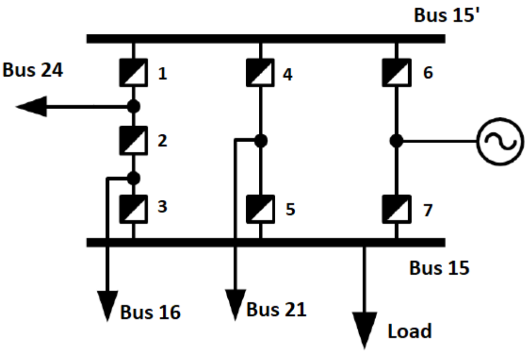

Figure 5.15: Scheme of internal topology breakers of the substation located on bus 15 with connections to respective buses.

Then, to substation 15 was applied 7 CNN models B, one for each breaker to classify their connectivity, remembering that it is achievable due to independence between them. The utilisation of model B requires the input of just 16 measurements and to first tests was chosen the necessary power flow variables linked to each breaker. The division of global substation topology on single breakers problem allows the possibility of picking the most interesting measurements correlated to each one.

Diagrams 5.16 and 5.17, shows the different correspondence of top 16 available power flow results between different breakers from the substation. The analysis with the probabilistic developed concept approves the scheme of switchers 5.15, where most informative measurements represent the near bus that the breaker can connect as well as the power flows that it can express. Comparing breaker 1 5.16 and breaker 3 5.17, it is possible observes that variables link bus 15 to bus 16 and beyond it, are most relevant to breaker 3 than breaker 1. Contrarily, the measurements associated with bus 24 and surround it are most significant to breaker 1 than 3.

Table 5.9: Application of Model B to breakers from substation 15 and accuracy of each one as the global efficiency of procedure.

Breaker	**1**	**2**	**3**	**4**	**5**	**6**	**7**	
no. of failed tests	0	4	10	2	0	0	1	
Epoch of best model	3	64	287	10	9	4	2	
Failed tests %	0.000%	0.133%	0.333%	0.067%	0.000%	0.000%	0.033%	**Substation topology**
Accuracy	100.00%	99.87%	99.67%	99.93%	100.00%	100.00%	99.97%	=99.40%

Such exhaustive previous study makes sense on a complex topology problem and to this substation of 7 breakers signify the use of 32 different power flow measurements due to the specific characteristics that each switcher affords.

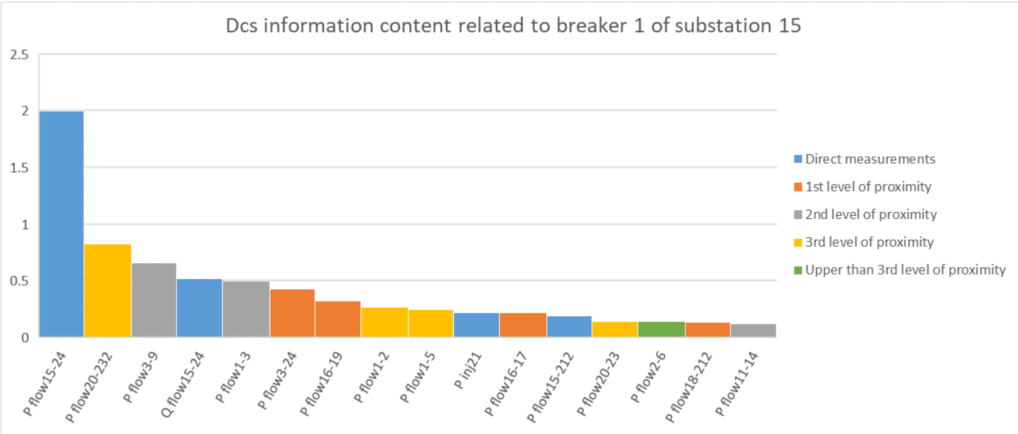

Figure 5.16: Breaker 1 from substation 15 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from substation localisation.

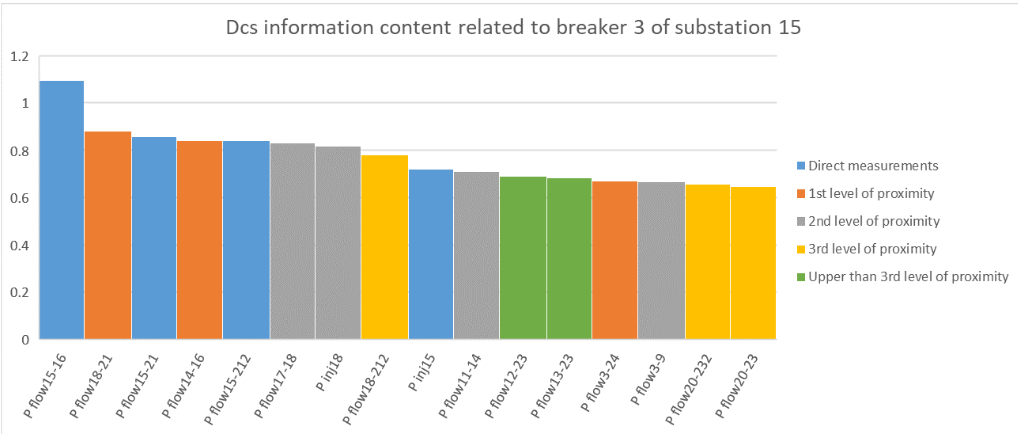

Figure 5.17: Breaker 3 from substation 15 related to the 16 most significant power flow variables ordered decreasingly, where measurements are represented besides 1^{st}, 2^{nd} and 3^{rd} proximity levels from substation localisation.

The efficiency of the proposed test is presented in table 5.9 as it is possible to see the accuracy of the classification procedure is almost exact, contributing to an efficiently global substation topology determination.

Same methodology ideas were applied to a more complex substation 5.18 with 13 incorporated breakers. For the present problem, switchers 11 and 12 that manage the possibility of connecting the load to bus 9 are considered as always closed, the focus of work resides on the influence of the measurements on lines arrangement. Study of probabilistic information content was done to substation 9, has been common before all demonstrated tests. However, comparing to substation 15, this problem requires more different measurements, 45 contrasting to previous 32 measurements. The increment of 4 additional breakers explains increases of variables needed, so 45 measurements to substation 9 topology reconstruction are adequate to the complexity of the problem.

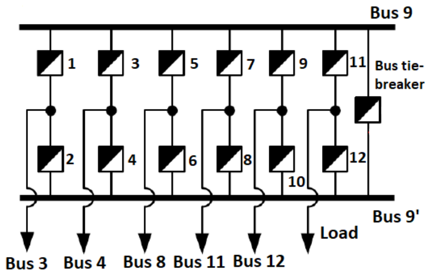

Figure 5.18: Scheme of internal topology breakers of the substation located on bus 9 with connections to respective buses.

Thus, accuracy results from proposed idea are presented in table 5.10. As predictable, a substation with more breakers and consequently with a higher number of complex arrangements determine a lower global efficiency comparing to a substation with less incorporated switchers. Comparing the training procedure of each one, was necessary a training execution with a vast number of epochs in some breakers than others but globally, the $91,95\%$ of accuracy determines a satisfactory classification precision.

Table 5.10: Application of Model B to breakers from substation 9 and the accuracy of each one as the global efficiency of the procedure.

Breaker	1	2	3	4	5	6	7	8	9	10	Bus tie-breaker	
no. of failed tests	17	11	17	76	4	45	1	39	1	39	0	
Epoch of best model	114	150	26	165	70	275	119	149	96	104	3	
Failed tests %	0.567%	0.367%	0.567%	2.533%	0.133%	1.500%	0.033%	1.300%	0.033%	1.300%	0.000%	Substation topology
Accuracy	99.43%	99.63%	99.43%	97.47%	99.87%	98.50%	99.97%	98.70%	99.97%	98.70%	100.0%	=91.95%

5.4 PMU introduction

Performed tests to breaker status classification and topology substation reconstruction considers only the existence of power flow measurements and power injections as well. On practical operation of the network, the system operator is connected to the data acquisition system where this type values are the most common even if few of them are available.

However, other variables of the system can be presented such as the bus voltage - magnitude and phase - that are available where a Phasor Measurement Unit (PMU) is installed on the bus where is necessary the report of this relevant information. Inspired by this idea, the present section of work is addressed to the possibility of PMU installation and how much the existence of these measurements allows the upgrade of substation topology determination accuracy.

Study of correlation between the voltage variables and the connectivity of internal breakers was done by the probabilistic distance between the open and close pdf of each topology variable. Naturally that the study was done to each switcher, although it is necessary to choose the pair voltage magnitude and phase that most contribute to increment of informative zone linked to study substation. So that, to substation 9 and 15, was tested the introduction of such information to previous power flow measurements and evaluated the two most candidates to perform a more efficient accuracy.

Firstly, to substation 15 was observed that voltage content most linked to the status of substation incorporated breakers are present on bus 15 and 24, such appreciation was observable made to determine a *à priori* selection. Thus, to test those same ideas, two global experiments were done. Firstly, the introduction of PMU measurements on each breaker that 100% efficiency was not attained the voltage measurements of bus 15, and replacing the two less power flow variables referent to each switcher. Results for this experiment are presented in the following tables:

Table 5.11: CNN model B classification to incorporated switchers on substation 15 with introduction of bus 15 voltage measurements.

Breaker	1	2	3	4	5	6	7	
no. of failed tests	0	3	12	0	0	0	1	
Epoch of best model	4	233	32	51	9	42	2	
Failed tests %	0.000%	0.100%	0.400%	0.000%	0.000%	0.000%	0.033%	Substation topology
Accuracy	100.00%	99.90%	99.60%	100.00%	100.00%	100.00%	99.97%	=99.43%

Table 5.12: CNN model B classification to incorporated switchers on substation 15 with introduction of bus 24 voltage measurements.

Breaker	1	2	3	4	5	6	7	
no. of failed tests	0	2	6	0	0	0	0	
Epoch of best model	3	126	360	6	9	4	1	
Failed tests %	0.000%	0.067%	0.200%	0.000%	0.000%	0.000%	0.000%	Substation topology
Accuracy	100.00%	99.93%	99.80%	100.00%	100.00%	100.00%	100.00%	=99.70%

For the present case study, the accuracy without the introduction of voltage information content is near from 100%, so are not expected a significant change with possible PMU existence. However, it is interesting to understand that voltage values from bus 24 are more valuable than voltage content resides on bus 15 - where the substation is localised. Going on a more in-depth analysis, the introduction of voltage from bus 15 does not affect the global accuracy of substation topology classification contrasting to the first test for this case study 5.9. Contrarily, voltage values from bus 24 upgrade the efficiency of topology determination with training procedure effort remaining acceptable in terms of run time.

Hence, this same made task for substation 15, was done to substation 9. For this substation, informative content analysis was done, and for all voltages of the test system, that probabilistic study determines informative local importance where bus 9 and the secondary of the substation, bus 25, are located. The results of this same experiences are presented in the tables 5.13 and 5.14.

Table 5.13: CNN model B classification to incorporated switchers on substation 9 with introduction of bus 9 voltage measurements.

Breaker	1	2	3	4	5	6	7	8	9	10	Bus tie-breaker	
no. of failed tests	34	14	23	70	4	24	1	45	1	48	0	
Epoch of best model	94	262	130	244	123	261	109	276	112	75	1	
Failed tests %	1.133%	0.467%	0.767%	2.333%	0.133%	0.800%	0.033%	1.500%	0.033%	1.600%	0.000%	Substation topology
Accuracy	98.87%	99.53%	99.23%	97.67%	99.87%	99.20%	99.97%	98.50%	99.97%	98.40%	100.00%	=91.52%

Table 5.14: CNN model B classification to incorporated switchers on substation 9 with introduction of bus 25 (secondary of substation 9) voltage measurements.

Breaker	1	2	3	4	5	6	7	8	9	10	Bus tie-breaker	
no. of failed tests	38	5	16	7	2	3	1	14	1	12	0	
Epoch of best model	188	28	95	221	25	34	107	329	161	425	1	
Failed tests %	1.267%	0.167%	0.533%	0.233%	0.067%	0.100%	0.033%	0.467%	0.033%	0.400%	0.000%	Substation topology
Accuracy	98.73%	99.83%	99.47%	99.77%	99.93%	99.90%	99.97%	99.53%	99.97%	99.60%	100.00%	=96.74%

For the substation 9, the first group of tests 5.10 performed a satisfactory global classification precision, although with significant space to improve the attained model to a best one. Differences between test 5.13 and 5.14 are notorious, one the first with the introduction of voltage information from bus 9, the efficiency decrease comparing to the initial test. For the first test, the recognition of inserted pattern on CNN model B defines a problematic learning task to this neural network with more epochs needed to achieve an adequate model.

Contrarily, the introduction of voltage measurements from secondary of bus 9 substation adds relevant information to classify each breaker correctly, principally where previous probabilistic study determines the influence of those variables on switcher connectivity. Breakers 1, with the addition of voltage magnitude and phase, reveals the worst performance compared to the initial test. To this breaker, without voltage variables were possible with 16 power flow results to define patterns correctly to a CNN better execution. Differently, to breaker 9 is clear the redundancy of voltage variables introduction as inputs of the CNN model, without introduces any considerable change.

Table 5.15: CNN model B classification to incorporated switchers on substation 9 with introduction of best results of bus 25 (secondary of substation 9) voltage measurements experience.

Breaker	1	2	3	4	5	6	7	8	9	10	Bus tie-breaker	
V_25	without	with	with	with	with	with	with	with	without	with	without	
no. of failed tests	17	5	16	7	2	3	1	14	1	12	0	
Epoch of best model	114	28	95	221	25	34	107	329	96	425	1	
Failed tests %	0.567%	0.167%	0.533%	0.233%	0.067%	0.100%	0.033%	0.467%	0.033%	0.400%	0.000%	Substation topology
Accuracy	99.43%	99.83%	99.47%	99.77%	99.93%	99.90%	99.97%	99.53%	99.97%	99.60%	100.00%	=97.43%

As expected, the introduction of a PMU does not affect all the breakers positively, but the choice of the most relevant PMU localisation can bring valuable contributions to the improvement of global classification accuracy. Such concept is evident in the last example, and the combination of breakers performances where voltage from secondary of substation 9 are present and cases where not, introduce considerable gains and are available on table 5.15.

Introduction of metering tools on that point insert an efficiency upgrade of $5,48\%$ in the same conditions of first performed test with variables reduction. Thus, developed work presented in this section proves the correlation between breakers connectivity and voltage measurements. Naturally, it was expected that this topology concern is directly linked to line power flows. Also, the voltage variables can contribute significantly to improve informative zones surrounding substation localisation.

5.5 Final Remarks

The present chapter proves the definition of a topology processor based on CNN as an efficient tool to classify various network topology problems. However, as mentioned before, what characterises a topology processor is the efficiency but also the time duration of training and classification procedure.

Along with the chapter results, it is shown the epochs number of the best model attained to each switch mode classification. However, the reference of it only serves as a term of comparison between them and, for example, a training procedure with hundreds of iterations does not mean that it takes to much time in practice. The time duration is dependent on the computational tools where the code is run, and with proper tools that means few seconds to classify breaker connectivity.

Chapter 6

Conclusions and Future Work

The present chapter emerges from the necessity of summarising how the developed work contributes to improve the network topology determination. Thus, the first section is addressed to the conclusions of this dissertation and the second to what there is left to improve on CNN modelling as a topology processor.

6.1 Conclusions

The monitoring of networks by the system operator is a concern that follows it's management until nowadays. This same issue guides the development of the present work, namely, with the definition of a unique topology processor model that is much more accurate on a realistic operation scenario approach.

However, the definition of a topology processor was defined by many steps. First of all, and without taking into account the constructed tool, a question follows its definition. What measurements are correlated to the breaker status determination? An answer was found on ITL techniques, mainly, on the probabilistic distance of Cauchy-Schwarz D_{CS}. Also dependent on this probabilistic tool was the Parzen Windows method of the pdf reconstruction, establishing a correlation between power flow past data and the connectivity of a breaker. This study of measurements correlation to the switch mode is the basis of all performed tests, giving an associated ranking of data and awareness of power flow behaviour through the lines.

Followed by the paradigm change and introduction of Deep Learning techniques on electrical engineering problems, especially, on the topology problem issue, and the focus goes to CNN framework. Most of the present document is addressed to the corroboration of CNN as a properly topology processor.

Many aspects of the presented models were tested, and the main conclusions are: the number of free parameters is directly linked to the effort optimisation of it; the standardisation of input data defines an adequate training procedure and, the ReLU emerges as the best activation function of the hidden layer. Other aspects as the size of filters that compose convolutional layers or the model depth are not related to the accuracy of the CNN model.

Otherwise, the proposed input values arrangement reveals a good strategy in order to present to the CNN model a better pattern definition and consequently, an easily trained model and a better accuracy acquisition. The increment of accuracy that results from this idea also proves the ITL developed informative tool as a correct approach in order to define a ranking of measurements.

The constructed neural network model performs a better breakers classification, even on an unobservable system. The lack of information affects the results of some breakers, but in most of them, the precision of the classification task resides almost unaffected. However, the test of this tool on a realistic operation scenario of 21,05% awareness of power flow measurements, defines CNN as a proper topology processor.

Focusing on the substation topology determination, the classification task on the studied substations from a group of CNN models defines a satisfactory result. Moreover, it was possible to observe the degree of complexity of this second problem on an operation scenario with few measurements. The main contribution of the developed strategy was the problem approach by a single breaker classification task of each incorporated switcher. So, it means that the number of utilised neural network models is equal to the number of breakers, contrarily to the developed works until today. The developed topology processor could be applied to network operation centres and help the system operator on daily control decisions. This tool could also supply different control automatisms, mainly, on a gradual transition to establish a self-healing network without human intervention.

Another relevant contribution of this dissertation is the ability to plan the meter installation. The future of the network operations will be guided by the developed tools, as what was demonstrated by this work. Thus, the ITL techniques prove their contributions to define informative zones in order to feed such models. For this particular problem, it was proved what the optimal localisation of power flow meters was, as well as PMU to improve the accuracy of the classification task.

Finally, this dissertation could inspire future works, using CNN as the main framework in order to extract patterns of values with some correlation between them. Since the same considerations that were taken into account in this dissertation also define the modelling of new problems. The CNN was usually applied to image recognition but also plays an essential role in data pattern acquisition.

6.2 Future Work

Although the developed topology processor performs an adequate classification of the breaker connectivity, some upgrades could be accomplished with future work. So, as all scientific work is in constant progress, the developed CNN model can be improved, in order to achieve a practical application with the following ideas:

- The improvement of the training procedure using other regularisation techniques as the dropout of neurons on hidden layer, the application of the L1 or L2 regularisation methodologies;

- Prove the real-time application of this model, using a GPU as an acceleration tool. Such hardware device would allow the time reduction of training procedure to a few seconds. A system operator with this topology processor could in real-time be efficiently aware of the real state of an uncovered part of the network;
- Testing different matrix input organisations. A new scheme of values arrangement could improve the classification accuracy;
- On a practical implementation, creates a data validation system. To preserve the existence of a unique dataset, it is necessary checking if an acquired measurement represents a new contribution to improve the CNN model training or instead of it, increases dataset redundancy;

Appendix A

Test System

Table A.1: Parameters of IEEE 24-bus test system.

bus i	type	V_n (kV)	P_D (MW)	Q_D (MVAr)	P_G (MW)	$Q_{Gm}in$ (MVAr)	$Q_{Gm}ax$ (MVAr)	B_s (MVAr at V = 1.0 p.u.)	V_sp (p.u.)
1	PV	138	108	22	172	-50	80	0	1.035
2	PV	138	97	20	172	-50	80	0	1.035
3	PQ	138	180	37	-	-	-	0	-
4	PQ	138	74	15	-	-	-	0	-
5	PQ	138	71	14	-	-	-	0	-
6	PQ	138	136	28	-	-	-	-100	-
7	PV	138	125	25	240	0	180	0	1.025
8	PQ	138	171	35	-	-	-	0	-
9	PQ	138	175	36	-	-	-	0	-
10	PQ	138	195	40	-	-	-	0	-
11	PQ	230	-	-	-	-	-	0	-
12	PQ	230	-	-	-	-	-	0	-
13	REF	230	265	54	285.3	0	240	0	1.02
14	PV	230	194	39	0	-50	200	0	0.98
15	PV	230	317	64	215	-50	110	0	1.014
16	PV	230	100	20	155	-50	80	0	1.017
17	PQ	230	-	-	-	-	-	0	-
18	PV	230	333	68	400	-50	200	0	1.05
19	PQ	230	181	37	-	-	-	0	-
20	PQ	230	128	26	-	-	-	0	-
21	PV	230	-	-	400	-50	200	0	1.05
22	PV	230	-	-	300	-60	96	0	1.05
23	PV	230	-	-	660	-125	310	0	1.05
24	PQ	230	-	-	-	-	-	0	-

References

[1] Gu Xinxin, Ning Jiang, and China Electric Power Press. *Self-healing Control Technology for Distribution Networks*. Wiley, 2017.

[2] A Monticelli. *State Estimation in Eletric Power System*. 1999.

[3] Fred C. Schweppe and D. Rom. Power System Static-State Estimation, Part II: Approximate Model. *IEEE Transactions on Power Apparatus and Systems*, PAS-89(1):125–130, 1970. doi:10.1109/TPAS.1970.292678.

[4] Fred C. Schweppe. Power System Static-State Estimation, Part III: Implementation, 1970. doi:10.1109/TPAS.1970.292680.

[5] F C Schweppe and J Wildes. Power System Static-State Estimation, Part I: Exact Model, 1970. doi:10.1109/TPAS.1970.292678.

[6] Ali Abur and Antonio Gómez Expósito. *Power System State Estimation: Theory and Implementation*. 2004.

[7] R. L. Lugtu, D. F. Hackett, K. C. Liu, and D. D. Might. Power system state estimation: Detection of topological errors. *IEEE Transactions on Power Apparatus and Systems*, PAS-99(6):2406–2412, 1980. doi:10.1109/TPAS.1980.319807.

[8] K. A. Clements and P. W. Davis. Detection and identification of topology errors in electric power systems. *IEEE Transactions on Power Systems*, 3(4):1748–1753, Nov 1988. doi:10.1109/59.192991.

[9] F. F. Wu and W. . E. Liu. Detection of topology errors by state estimation (power systems). *IEEE Transactions on Power Systems*, 4(1):176–183, Feb 1989. doi:10.1109/59.32475.

[10] A. Simoes Costa and J. A. Leao. Identification of topology errors in power system state estimation. *IEEE Transactions on Power Systems*, 8(4):1531–1538, 1993. doi:10.1109/59.260956.

[11] M.R. Irving and M.J.H. Sterling. Substation data validation. *IEE Proceedings C Generation, Transmission and Distribution*, 129(3):119, 2010. doi:10.1049/ip-c.1982.0018.

[12] N. Singh and F. Oesch. Practical experience with rule-based on-line topology error detection. *IEEE Transactions on Power Systems*, 9(2):841–847, May 1994. doi:10.1109/59.317631.

[13] A. Monticelli. Modeling circuit breakers in weighted least squares state estimation. *IEEE Transactions on Power Systems*, 8(3):1143–1149, 1993. doi:10.1109/59.260883.

[14] A. Monticelli and A. Garcia. Modeling zero impedance branches in power system state estimation. *IEEE Transactions on Power Systems*, 6(4):1561–1570, Nov 1991. doi:10.1109/59.117003.

[15] O. Alsaç, N. Vempati, B. Stott, and A. Monticelli. Generalized state estimation. *IEEE Transactions on Power Systems*, 13(3):1069–1075, 1998. doi:10.1109/59.709101.

[16] P. D. Yehsakul and I. Dabbaghchi. A topology-based algorithm for tracking network connectivity. *IEEE Transactions on Power Systems*, 10(1):339–346, Feb 1995. doi:10.1109/59.373954.

[17] Ali Abur, Mehmet Aelik, and Hongrae Kim. Identifying the Unknown Circuit Breaker Statuses in Power Networks. *IEEE Transactions on Power Systems*, 10(4):2029–2037, 1995. doi:10.1109/59.476072.

[18] H. Singh and F. L. Alvarado. Network topology determination using least absolute value state estimation. *IEEE Transactions on Power Systems*, 10(3):1159–1165, Aug 1995. doi:10.1109/59.466541.

[19] L. Mili and G. Steeno. A robust estimation method for topology error identification. *IEEE Transactions on Power Systems*, 14(4):1469–1476, 1999. doi:10.1109/59.801932.

[20] K. A. Clements and A. S. Costa. Topology error identification using normalized lagrange multipliers. *IEEE Transactions on Power Systems*, 13(2):347–353, May 1998. doi:10.1109/59.667350.

[21] J. Pereira, V. Miranda, and J. T. Saraiva. A comprehensive state estimation approach for ems/dms applications. In *PowerTech Budapest 99. Abstract Records. (Cat. No.99EX376)*, pages 272–, Aug 1999. doi:10.1109/PTC.1999.826704.

[22] Jorge Pereira, Vladimiro Miranda, and J.T. Sataiva. Combining Fuzzy and Probabilistic Data in Power System State Estimation. *Proceedings of PMAPS'97 - Probabilistic Methods Applied to Power Systems*, pages 151–157, 1997.

[23] Jorge Pereira and Vladimiro Miranda. Fuzzy control of state estimation robustness. (June):24–28, 2002.

[24] Jorge Pereira. *A State Estimation Approach for Distribution Networks Considering Uncertainties and Switching*. PhD thesis, 2001.

[25] T. V. Cutsem, M. Ribbens-Pavella, and L. Mili. Hypothesis testing identification: A new method for bad data analysis in power system state estimation. *IEEE Transactions on Power Apparatus and Systems*, PAS-103(11):3239–3252, Nov 1984. doi:10.1109/TPAS.1984.318561.

[26] Elizete Maria Lourenço, Antonio Simões Costa, and Kevin A. Clements. Bayesian-based hypothesis testing for topology error identification in generalized state estimation. *IEEE Transactions on Power Systems*, 19(2):1206–1215, 2004. doi:10.1109/TPWRS.2003.821442.

[27] Elizete Maria Lourenço, Kevin A. Clements, and Antonio Simões Costa. A topology error identification method directly based on collinearity tests. *IEEE Russia PowerTech*, pages 1–6, 2005.

[28] Eduardo Caro, Antonio J. Conejo, and Ali Abur. Breaker status identification. *IEEE Transactions on Power Systems*, 25(2):694–702, 2010. doi:10.1109/TPWRS.2009.2035321.

[29] A. P. Alves da Silva, V. H. Quintana, and G. K. H. Pang. Solving data acquisition and processing problems in power systems using a pattern analysis approach. *IEE Proceedings C - Generation, Transmission and Distribution*, 138(4):365–376, July 1991. doi:10.1049/ip-c.1991.0046.

[30] A. P. Alves da Silva, V. H. Quintana, and G. K. H. Pang. Neural networks for topology determination of power systems. In *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, pages 297–301, July 1991. doi:10.1109/ANN.1991.213459.

[31] A. P. Alves da Silva, V. H. Quintana, and G. K. H. Pang. A pattern analysis approach for topology determination, bad data correction and missing measurement estimation in power systems. In *Proceedings of the Twenty-Second Annual North American Power Symposium*, pages 363–372, Oct 1990. doi:10.1109/NAPS.1990.151390.

[32] D. M. Vinod Kumar, S. C. Srivastava, S. Shah, and S. Mathur. Topology processing and static state estimation using artificial neural networks. *IEE Proceedings - Generation, Transmission and Distribution*, 143(1):99–105, Jan 1996. doi:10.1049/ip-gtd:19960050.

[33] Jakov Krstulovic, Vladimiro Miranda, Antonio J.A. Simoes Costa, and Jorge Pereira. Towards an auto-associative topology state estimator. *IEEE Transactions on Power Systems*, 28(3):3311–3318, 2013. doi:10.1109/TPWRS.2012.2236656.

[34] Jakov Krstulović and Vladimiro Miranda. Denoising auto-associative measurement screening and repairing. *2015 18th International Conference on Intelligent System Application to Power Systems, ISAP 2015*, pages 1–6, 2015. doi:10.1109/ISAP.2015.7325548.

[35] J. Krstulović and V. Miranda. Selection of measurements in topology estimation with mutual information. In *2014 IEEE International Energy Conference (ENERGYCON)*, pages 589–596, May 2014. doi:10.1109/ENERGYCON.2014.6850486.

[36] Vladimiro Miranda, Jakov Krstulovic, Hrvoje Keko, Cristiano Moreira, and Jorge Pereira. Reconstructing missing data in state estimation with autoencoders. *IEEE Transactions on Power Systems*, 27(2):604–611, 2012. doi:10.1109/TPWRS.2011.2174810.

[37] Vladimiro Miranda, Jakov Krstulovic, Joana Hora, Vera Palma, and José C Príncipe. Breaker status uncovered by autoencoders under unsupervised maximum mutual information training. pages 1–6, 2013.

[38] J K Opara. *Information Theoretic State Estimation in Power Systems*. PhD thesis, University of Porto, Portugal, 2013.

[39] Ian Goodfellow Courville, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016. URL: http://www.deeplearningbook.org.

[40] Cyxtera Technologies. Building AI Applications Using Deep Learning. URL: https://blog.easysol.net/building-ai-applications/.

[41] Yann LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*. arXiv:1004.3732, doi:10.1162/neco.1989.1.4.541.

[42] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998. doi:10.1109/5.726791.

[43] Karpathy Andrej. CS231n Convolutional Neural Networks for Visual Recognition. URL: http://cs231n.github.io/convolutional-networks/.

[44] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew Back. Face Recognition: A Convolutional Neural Network Approach. *Neural Networks, IEEE Transactions*, 8:98 – 113, 1997. doi:10.1109/72.554195.

[45] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. pages 815–823, 2015. doi:10.1109/CVPR.2015.7298682.

[46] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014. arXiv:1501.05703, doi:10.1109/CVPR.2014.220.

[47] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-Scale Video Classification with Convolutional Neural Networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6909619, arXiv:1412.0767, doi:10.1109/CVPR.2014.223.

[48] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter:4489–4497, 2015. arXiv:1412.0767, doi:10.1109/ICCV.2015.510.

[49] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3D Convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):221–231, 2013. arXiv:1102.0183, doi:10.1109/TPAMI.2012.59.

[50] Gul Varol, Ivan Laptev, and Cordelia Schmid. Long-Term Temporal Convolutions for Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1510–1517, 2018. arXiv:1604.04494, doi:10.1109/TPAMI.2017.2712608.

[51] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014. arXiv:1311.2524, doi:10.1109/CVPR.2014.81.

[52] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017. arXiv:1506.01497, doi:10.1109/TPAMI.2016.2577031.

[53] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and*

Pattern Recognition, 07-12-June:1–9, 2015. arXiv:1409.4842, doi:10.1109/CVPR.2015.7298594.

[54] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. arXiv:1606.00915, doi:10.1109/TPAMI.2017.2699184.

[55] Mohsen Guizani, Di Wu, Min Chen, Xiaobo Shi, and Yin Zhang. Deep Features Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network. *IEEE Transactions on Big Data*, 7790(c):1, 2017. doi:10.1109/tbdata.2017.2717439.

[56] Mufti Mahmud, Mohammed Shamim Kaiser, Amir Hussain, and Stefano Vassanelli. Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(6):2063–2079, 2018. arXiv:1711.03985, doi:10.1109/TNNLS.2018.2790388.

[57] Vladimiro Miranda, Pedro A. Cardoso, Ricardo J Bessa, and Ildemar Decker. Through the looking glass: Seeing events in power systems dynamics. *International Journal of Electrical Power and Energy Systems*, 106(October 2018):411–419, 2019. URL: https://doi.org/10.1016/j.ijepes.2018.10.024, doi:10.1016/j.ijepes.2018.10.024.

[58] C.E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(July 1928):379–423, 1948.

[59] E. Parzen. *On estimation of a probability density function and the mode*, volume 37. 1951. arXiv:arXiv:1011.1669v3, doi:10.1214/aoms/1177705148.

[60] V Miranda. "information theoretic learning principles a short tutorial". In *ISAP Conference and debate*, pages 12–15, 2015.