Universidad Técnica Federico Santa María
Departamento de Ingeniería Eléctrica
Santiago de Chile

Machine Learning Aided Column Generation for Solving Transmission Network Expansion Planning

Pablo Felipe Oteíza Canales

Propuesta de tesis para optar al grado de
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA ELÉCTRICA

y al título profesional de
INGENIERO CIVIL ELECTRICISTA

Profesor guía

RICARDO ÁLVAREZ

Profesor co-guía

Mehrdad Pirnia

18 de marzo de 2024

Propuesta de tesis para optar al grado de
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA ELÉCTRICA
y al título profesional de
INGENIERO CIVIL ELECTRICISTA

Machine learning aided column generation for solving transmission network expansion planning.

Candidato
Pablo Felipe Oteíza Canales

Universidad Técnica Federico Santa María
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

Profesor guía
RICARDO ÁLVAREZ

Profesor co-guía
Mehrdad Pirna

Santiago, 03 de diciembre de 2024

Table of contents

Figure index.

1. Introduction

1.1. Motivation

Many countries around the world are experiencing an energy transition towards low-carbon economies, in order to reduce greenhouse gas emissions and combat climate change. One of the main sectors called to lead this transition is the electricity sector, mainly through the massive incorporation of renewable energies and the decommission of coal-fired power plants. In Chile, according to the long-term Chilean energy policy, the target is to reach net zero emissions by 2040 and supply 100% of the electrical demand from renewable energies by 2050 [1]. The Chilean commitment to combat climate change already started in 2011 by promoting the introduction of renewable energies in the electrical sector. Since then, the installed capacity of variable renewable energies (VRE) increased from 4% of the total capacity in 2011, to more than 40% in 2023.

Under this global trend, the transmission infrastructure is key for achieving a cost-effective and secure energy transition. Indeed, to have an adequate transmission capacity is fundamental to accommodate new renewables and facilitate their integration into the market, supply the growing demand in a cost-effective way and enhance competition to ensure market efficiency [2]. Increasing transmission capacity is not an easy task, due to growing difficulties in acquiring new rights-of-way, the opposition of local communities to the construction of new lines and long processes of environmental licensing, among others [3]. The aforementioned challenges considerably extend the development time of new transmission line projects. Delays in the commission of expansion projects can have significant technical and economic repercussions. Key concerns include unexpected curtailment of renewable energy sources (RES), where new RES investments can experience unexpected, diminished returns if the transmission capacity if not available on time for the delivery of their energy into the grid, increasing energy costs for consumers and greater transmission losses caused by the sub optimal operation of the transmission network due to underdevelopment. Flexible technological solutions can alleviate these issues by enhancing the network capacity and reliability in shorter time and without the need of installing new transmission lines. Amongst these alternative technological solutions, we can find line reconductoring, HVAC to HVDC reconversion, Flexible AC Transmission Systems (FACTS), Battery Energy Storage System (BESS) installed as grid boosters and others. What all these solutions have in common is shorter development times, lower investment cost and less challenging regulatory processes compared to the investment in new transmission lines. Consequently, nowadays there are plenty of technological options to increase the transmission capacity. However, the main challenge consists of determining which technology, when and where should be incorporated into the system to minimize total system costs while maintaining security, social and/or environmental constraints. To answer this question, a usual approach is to address this challenge as an optimization problem, the transmission network expansion planning (TNEP) problem.

The TNEP is usually formulated as a mixed-integer linear problem (MILP) [4], where integer variables are related to the investment decisions, while continuous variables are related to the power system operation. In more complex formulations, integer variables can also comprise operational variables such as the generator's unit commitment [5]. The TNEP can also be formulated as a nonlinear problem (MINLP), for example to consider power system losses and/or AC power flows [6]. Furthermore, the use of stochastic and/or robust models has increased in the last years, due to the growing uncertainty involved in the network planning. In the past, when power systems were vertically integrated and the generation matrix was dominated by conventional fossil-fuel based generation, there was little uncertainty regarding future generation capacity and power feed-in [7]. Therefore, the majority of TNEP formulations were deterministic. However, the deregulation of the electricity markets, along with the massive introduction of renewable energies, brought new uncertainties regarding future generation capacity and availability. The level of uncertainty regarding future generation capacity increased even more with the introduction of renewable energies, because of their short construction lead times (1 – 2 years), compared to conventional fossil-fuel based generation (3 – 5 years), thus increasing the need of adopting stochastic or robust TNEP formulations. Furthermore, it is expected that the use of stochastic and robust model will increase in the future, in order to incorporate new sources of uncertainties, for examples, those related to long-term effects of the climate change.

In all cases, the TNEP problem is NP-hard, which means that it cannot be solved in polynomial time. For large-scale realistic-size power system models, solving the TNEP problem demands significant amounts of computational resources and could even require several days to be solved using a state-of-the-art optimization algorithm. Furthermore, the appearance of emergent technologies that increase the operational flexibility of power systems have introduced new challenges for the TNEP. Examples of such technologies are FACTS [8], storage devices [9], special protection schemes [10], energy storage-based grid booster [11] and HVDC links [12], as well technological option to increase the use of existing transmission network assets such as line uprating [13] and HVDC conversion [14]. For each new technology, novel formulations capable of capturing their value in the power system are required, which usually means having more complex formulations, more variables, and constraints, thus resulting in models which are even more challenging to solve. In particular, proper models of energy storage systems require the introduction of time-coupling constraints to account for the energy balance, thus significantly increasing the complexity of the TNEP problem. Note that, to capture the benefits of short-term storage systems such as batteries, requires increasing the time resolution of the models (for example, representative days instead of isolated operating conditions) [15], which in turn increase the computational burden of the TNEP formulation.

To make the TNEP problem more tractable, it is usual to adopt simplifications such as i) using reduced network representations, ii) limiting the planning horizon (for multi-year TNEP), iii) limiting the number of representative operational conditions and iv) neglecting operational constraints, for example, those related to the generator's unit commitment. Although simplifying power system models is necessary to obtain solutions in reasonable time using available computational resources, too many simplifications can limit the practical application of the solutions. Consequently, there is a growing need of developing new and more advanced optimization algorithms, in order to address more complex and accurate TNEP problem formulations. In this regard, one strategy that has received increasing attention in the last years is the use of Artificial Intelligence (AI) techniques for solving complex optimization problems. AI techniques, as a universal function approximator, have shown great potential as an aid to solve optimization problems [16]. It has been applied to areas such as healthcare [17], finance [18], logistics [19], energy management [20], and resource allocation [21]. The ability of machine learning to recognize patterns and identify efficient policies can also help addressing complex problems. Despite of this, the development of globally optimal algorithms that integrate AI techniques with mathematical optimization ones is still incipient in power system planning problems, such as the TNEP, which is the main subject of this thesis.

1.2. Hypothesis

The main hypothesis of this work is that AI techniques can be combined with decomposition-based traditional mathematical optimization ones to reduce the time required to solve large-scale MILP problems guaranteeing, at the same time, the optimality of the solution.

1.3. Objectives

1.3.1. Main objective

The main objective of this thesis is to propose a novel optimization algorithm that combines the use of artificial intelligence techniques with Column Generation, to solve the TNEP problem. The proposed algorithm must ensure the optimality of the solutions found and require less computational time to converge, compared to the traditional Column Generation algorithm.

1.3.2. Specific objectives

- To review the state of the art on artificial intelligence techniques applied to solve large scale optimization problems, with a special focus on power system planning and mixed integer optimization problems.
- To develop case studies of the TNEP problem with different levels of details. While small case studies will be used for developing purposes, large-scale case studies will be used to evaluate the performance of the proposed optimization algorithm.
- To propose a novel optimization algorithm that can accelerate the convergence speed of the column generation algorithm with the assistance of machine learning without compromising optimality.

- To evaluate the performance of the proposed optimization algorithm in solving the TNEP problem for a case study based on the Main Chilean Power System (SEN). The performance will be evaluated in terms of quality of the solution and computational time and compare it against the ones obtained using the algorithm Column Generation.

2. Algorithms for solving the TNEP problem

2.1. The TNEP problem

The Transmission Network Expansion Planning (TNEP) consists of determining the optimal investment decision (what, where and when) in the transmission network to supply the projected demand at minimum cost, fulfilling at the same time technical, economical and or environmental constraints [4]. Models in TNEP differentiate from each other on several points. According to the planning horizon, the TNEP problem can be divided into short-term (1 to 5 years), mid-term (5 to 10 years) and long-term planning (> 10 years) [22]. According to the treatment of uncertainties, TNEP models can be deterministic [23] [24] [25] [26], stochastic [27] [28] or robust [29] [30]. Typical uncertainties in the TNEP are the demand growth, fuel prices and availability, expansion of the generation park and hydrology. While deterministic TNEP formulations assume one possible scenario realization, stochastic and robust models consider several possible outcomes of the uncertain variables, which are usually characterized through scenarios. According to the operational constraints, TNEP formulations can use DC or AC power flow, or a mixture of thereof. On the one hand, DC formulations [31] [32] only consider active power flows and assume equal voltage magnitude on all buses [33]. Here, it is also usual to neglect the power line losses and other sources of non-linearities, resulting in mixed-integer linear problems (MILP). Most TNEP fall into this category. On the other hand, AC formulations consider both, active and reactive power flows, voltage magnitude and angles in all busbars and power losses, among others. Although an AC formulation is a more accurate representation of the power system operation, the formulation has non-linearities and is therefore much harder to solve, compared to its DC-based counterpart [34]. According to the time-resolution for representing the yearly operating, TNEP models can use isolated operating conditions [35] or representative days [36]. In both cases, to keep the problem tractable.

For increasing the transmission capacity, the traditional approach has been to build new lines and transformers. However, the appearance of flexible technologies that allows increasing the operational flexibility of power systems has opened new opportunities in power system design. Consequently, more recent TNEP models have started to incorporate such technologies. Examples of these technologies are FACTS [8], storage devices [9], special protection schemes [10], energy storage-based grid booster [11] and HVDC links [12], as well as technological options to increase the use of existing transmission network assets such as line uprating [13] and HVDC conversion [14]. Works in [37] [38] [39] present different transmission planning models that incorporate these technologies into new formulations and showcase their value cost-effective alternative to increase transmission capacity and accommodate new RES generation. Nowadays, there is a wide consensus that considering flexible technologies in the TNEP not only results in more economic expansion plans, but also allows mitigating the impact of the uncertainties, defer investment decisions of new lines to later years and to alleviate congestions in the short run, where no traditional reinforcement is possible. In [40], a stochastic model for transmission and ESS expansion planning is presented. This work shows the capacity of ESS to alleviate the risks of uncertainties, especially due to their short deployment time (less than 1 year), compared to the time required to build new transmission lines, which can take several years. However, to properly capture the benefits of flexible technologies in transmission planning a greater time resolution is required for the operational model [41] [42], increasing computational cost for more detailed operational conditions. Considering new technologies also introduces additional constraints and decision variables into the formulation which drives up complexity of the model and results in greater computation cost.

2.2. Decomposition techniques on large scale optimization problems

As presented in the previous section solving the TNEP for realistic-sized power system is extremely challenging due to its NP-Hard nature. Hence, solving large-scale TNEP models in reasonable computational time requires advanced optimization techniques. The traditional multi-year TNEP model that will be used to highlight the decomposition techniques is the following:

$$min \sum_{y=1}^{Y} C_y^{INV} I_y + \sum_{y=1}^{Y} C_y^{op} X_y \tag{1}$$

Subject to:

$$Z_y \leq \sum_{r=1}^{y} I_r \quad \forall y \tag{2}$$

$$D_y X_y \leq Z_y \quad \forall y \tag{3}$$

$$X_y \in \chi_y \quad \forall y \tag{4}$$

The objective function in (1) consists of minimizing total system costs, represented as the sum of the investment costs and the operating costs. The investment decisions in each year y are represented by I_y and the operational variables by X_y. Constraints (2) relate the investment decisions I_y with the corresponding cumulated capacity Z_y, and constraints (3) restrict the power system operation according to the available capacity Z_y.

The TNEP problem formulation in (1) – (4) involves integer investment decisions (I_y) and continuous operating decisions (X_y). Since dealing with variables of integer nature is much harder than dealing with continuous ones, integer variables are called complicating variables. From (2) – (3) it can be seen that, if the investment variables are known and fixed, the power system operation can be solved independently for each year, because the cumulated capacity of each line is also known. However, constraints (2) prevent optimizing investment decisions for each year independently. Hence, these constraints are denoted as complicating constraints. Another important feature of the formulation (1) –(4) is that it has a block diagonal structure, as shown in Figure 1. While operational variables and restrictions are associated to one year only (constraint blocks for each year), decision investment variables in different years are linked together by complicating constraint blocks.

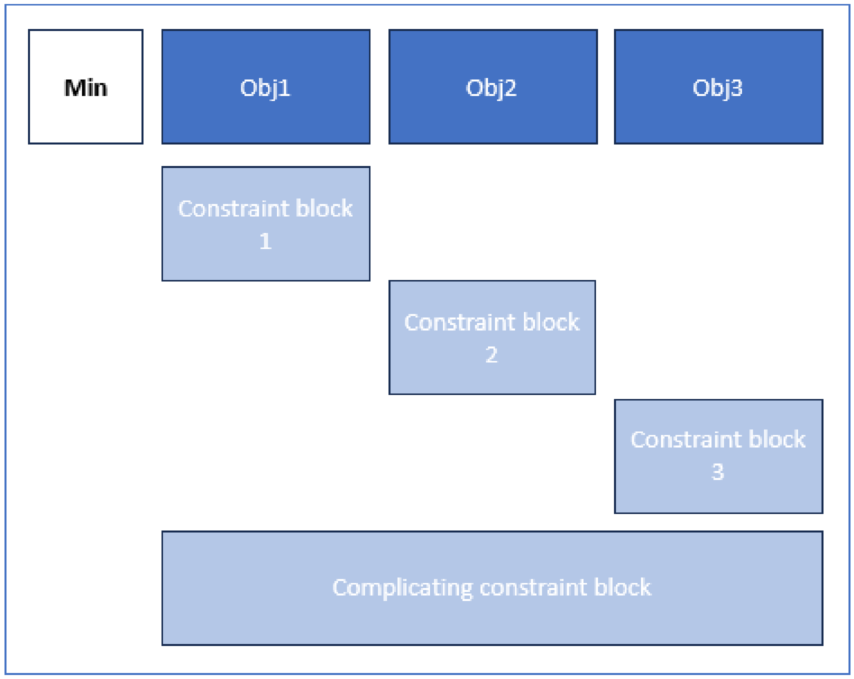

Figure 1 Illustration of a problem with block diagonal structure.

Decomposition techniques take advantage of the block structure and use the different nature of the decision variables to break down the original problem into smaller and more manageable sub problems. Using the solutions of these sub problems in conjunction with a coordinating master problem, decomposition techniques can iteratively approach the globally optimal solution of the original problem. The specific approach depends on the complicating variables, constraints, and the specific structure of the problem. In the TNEP problem, most traditional decomposition techniques are Bender's decomposition [43] [44] [45], Column Generation (CG) [46] [47] [48] and Lagrangian relaxation [49] [50] [51].

2.2.1 Benders decomposition

The Benders decomposition algorithm allows us to solve LP and MILP problems with complicating variables in a distributed manner using an iterative approach. The solution of the problem can be obtained by parametrizing it as a function of the complicating variables $(I_1, \ldots, I_n)$, related to the investment decisions. The master problem in Benders has the following structure:

$$min \sum_{y=1}^{Y} C_y^{INV} I_y + \sum_{y=1}^{Y} \alpha_y(I_1, \ldots, I_y) \qquad (5)$$

Subject to:

$$I_y \in \mathbb{Z}^{\mathbb{N}} \qquad \forall y \qquad (6)$$

Where:

$$\alpha_y(I_1, \ldots, I_y) = \min C_y^{op} X_y^{SP} \qquad \forall y \qquad (7)$$

Subject to:

$$Z_y = \sum_{r=1}^{y} I_r \quad \forall y \tag{8}$$

$$D_y X_y \leq Z_y \quad \forall y \tag{9}$$

$$X_y \in \chi_y \qquad \forall y \tag{10}$$

The advantage of the master problem is that it only has complicating variables, and this fact can be exploited computationally. However, to solve the master problem (5) – (7) an exact formulation of $\alpha_y(I_1,\ldots,I_y)$ is required, which is very difficult to obtain. Nonetheless, $\alpha_y(I_1,\ldots,I_y)$ is a convex function, as demonstrated in [52], and it can be approximated from below using hyperplanes, called Benders cuts. These cuts can be obtained iteratively by solving the following subproblem for each iteration k:

$$v_y^k = min\ C_y^{op} X_y \tag{11}$$

Subject to:

$$Z_y = \tilde{Z}_y^k: \quad \lambda_y^k, \quad \forall y \tag{12}$$

$$D_y X_y \leq Z_y \quad \forall y \tag{13}$$

$$X_y \in \chi_y \qquad \forall y \tag{14}$$

Where I_r^k is the solution obtained from the master problem in the iteration k and $\tilde{Z}_r^k = \sum_{r=1}^{y} I_r^k, \forall y$, represents the corresponding accumulated transmission capacity for year y. The variables λ_y^k are the optimal values of the dual variables associated with constraints (12) that fix the values of the complicating variables. As shown in [52], the function $\alpha_y(I_1,\ldots,I_y)$ can be approximated from below as follows:

$$\alpha_y(I_1,\ldots,I_y) \geq v_y^k + \lambda_y^k(Z_y - \tilde{Z}_y^k)$$

This lower bound of the function $\alpha_y(I_1,\ldots,I_y)$ allows us to formulate the following relaxed version of the master problem (5) –(6) for iteration i:

$$min \sum_{y=1}^{Y} C_y^{INV} I_y + \sum_{y=1}^{Y} \alpha_y \tag{15}$$

Subject to:

$$\alpha_y \geq v_y^k + \lambda_y^k(Z_y - \tilde{Z}_y^k), \forall y, \forall k \leq i \tag{16}$$

$$I_y \in \mathbb{Z}^{\mathbb{N}} \qquad \forall y \tag{17}$$

Since the Master problem is a relaxed version of the original problem, the value of the objective function in each iteration approximates the objective function of the original problem from below. Therefore, in each iteration, the optimal value of the objective function of the master problem (15) - (17) is a lower bound of the true solution of problem (5) - (6) for the same values of investment decision I_y. On the other hand, the subproblems are restricted versions of the original problem since the investment decisions are fixed. Consequently, in each iteration, the sum of the investment costs obtained in the master problem, which are suboptimal, and the operating costs obtained from the subproblems represents an upper bound of the optimal solution. It should be noted that, with increasing iterations, a better approximation of the cost function $\alpha_y(I_1,\ldots,I_y)$ is obtained, which makes the lower bound monotonously increasing. The algorithm stops when the gap between the upper and lower bounds are below a predefined objective gap size.

2.2.2 Largrangian Relaxation

Largrangian relaxation (LR) is a decomposition algorithm that can be applied to linear and nonlinear programing problems with complicating constraints. The basic idea of LR is to relax the complicated constrains by using Lagrange Multipliers, and then to decompose the problem into a series of sub problems for each stage. For the LR technique to be applied advantageously to a mathematical programming problem, the problem should have the following structure:

$$min\ f(x) \tag{18}$$

Subject to:

$$a(x) = 0 \tag{19}$$
$$b(x) \leq 0 \tag{20}$$
$$c(x) = 0 \tag{21}$$
$$d(x) \leq 0 \tag{22}$$

Where constraints (21) and (22) are complicating constraints, either because of the nature of the functions involved (e.g. nonlinear) or because of the presence of coupled variables that make the resolution of the original problem more difficult. Therefore, if these constraints are relaxed in some way, the problem can be drastically simplified.

To simply the problem, the LR technique decomposes the problem into a relaxed primal problem (RPP) that only depends on the non-complicating constraints and a dual problem (DP). The RPP is defined as follows:

$$\phi(\lambda, \mu) = \min_{x} L(x, \lambda, \mu) \tag{23}$$

Subject to:

$$a(x) = 0 \tag{24}$$
$$b(x) \leq 0 \tag{25}$$

Where the function $L(x, \lambda, \mu)$, known as the Lagrangian function, is defined as follows:

$$L(x, \lambda, \mu) = f(x) + \lambda^T c(x) + \mu^T d(x) \tag{26}$$

where λ and μ are the Lagrange multiplier vectors associated to the complicating constraints over $c(x)$ and $d(x)$, respectively. The DP has the following form:

$$\max_{\lambda,\mu} \phi(\lambda, \mu) \tag{27}$$

Subject to:

$$\mu \geq 0 \tag{28}$$

Where the dual function $\phi(\lambda, \mu)$ is concave and, in general, non-differentiable [52]. The LR is an attractive decomposition technique if the dual objective function $\phi(\lambda, \mu)$ can be easily evaluated for given values of the multiplier vectors, $\bar{\lambda}$ and $\bar{\mu}$. In other words, if it is easy to solve the RPP.

The RPP problem can usually be decomposed into a series of relaxed subproblems, thus facilitating its resolution. It also allows the exploitation of parallel computing. If the original problem is convex, the solution of the dual problem provides a solution of the primal problem as well. In the nonconvex case, the value of the objective function at the optimal solution of the dual problem provides a lower bound to the value of the objective function at the optimal solution of the primal problem. The difference between the optimal objective functions of the primal and dual problems is the duality gap.

For its use in the TNEP problem (1) - (4), the relaxed primal problem (RPP) using Lagrangian relaxation would have the following structure:

$$\phi(\lambda, \mu) = \min_{X,I} L(I, X, \lambda, \mu) \tag{29}$$

Subject to:

$$D_y X_y \leq Z_y \quad \forall y \tag{30}$$

$$X_y \in \chi_y \qquad \forall y \tag{31}$$

Where:

$$L(x, \lambda, \mu) = \sum_{y=1}^{Y} C_y^{INV} I_y + \sum_{y=1}^{Y} C_y^{op} X_y + \sum_{y=1}^{Y} \mu_y^T \cdot (Z_y - \sum_{r=1}^{y} I_r) \tag{32}$$

The associated dual problem has the same structure presented in (27) and (28). Note that, since the RPP only depends on non-complicating constraints, it can be solved for each year independently and, consequently, in parallel.

The problem is initialized with arbitrary values for the dual variables λ and μ. With these initial values, the RPP can be solved and obtain the optimal value of the operational *and* investment variables X, I respectively. In addition, the corresponding value of the objective function $\phi(\lambda, \mu)$ which represents a lower bound for the problem. Then in order to formulate the next iteration of the RPP the dual values must be updated according with the sub gradient rule using the following equations:

$$\lambda^{v+1} = \lambda^{v+1} + k^v \frac{c(x^v)}{\|c(x^v), d(x^v)\|} \tag{33}$$

$$\mu^{v+1} = \mu^{v+1} + k^v \frac{d(x^v)}{\|c(x^v), d(x^v)\|} \tag{34}$$

Where k is an arbitrary function that tends to zero with increasing number of iterations and the sum of k over all iterations tends to infinity. For example, the following function is usually used:

$$k^v = \frac{1}{a + bv} \tag{35}$$

Usually, the convergence criterion for stopping the algorithm is when changes in the dual values are marginal (bellow a threshold), for example when:

$$\frac{\lambda^{v+1}-\lambda^{v-1}}{\lambda^{v}} \leq \varepsilon \tag{36}$$

$$\frac{\mu^{v+1}-\mu^{v-1}}{\mu^{v}} \leq \varepsilon \tag{37}$$

If both criteria are meet the algorithm has achieved a sufficiently optimal solution. This process is repeated over several iterations until the desired tolerance ε has been achieved.

2.2.3 Column Generation

The column generation (CG) is an optimization technique suitable for solving problems with complicating constraints and block diagonal structure. If the variables of the problem are continuous, the algorithm is known as Dantzing Wolfe. If some of the variables are binary or integer (like in MILP problems) the algorithm is known as branch and price.

The CG algorithm solves the optimization problem by decomposing it into a reduced master problem (RMP) containing only complicating constraints and several subproblems. Basis of CG is that any feasible solution of the RMP can be written as a convex combination of the extreme points and directions of the polyhedron defining the feasible region [52]. This means that the optimal solution can also be written as a linear combination of the extreme points and directions. Theoretically, if all extreme points and direction of the polyhedron are known, the optimal solution can be obtained by simply solving the RMP. However, these extreme points and directions is very hard to obtain. Then, the CG algorithm consists of finding extreme points and directions iteratively, until the optimal solution is found.

Using the aforementioned concept, the CG algorithm decomposes the TNEP problem (1) - (4) into the following RMP:

$$\min_{\lambda,I} \sum_{y=1}^{Y} C_y^{inv} I_y + \sum_{y=1}^{Y}\sum_{i=1}^{k} C_y^{opt} X_y^i \lambda_y^i \tag{38}$$

Subject to:

$$\sum_{i=1}^{k} Z_y^i \lambda_y^i \leq \sum_{r=1}^{Y} I_r \ \forall y\colon \pi_y \tag{39}$$

$$\sum_{i=1}^{k} \lambda_y^i = 1 \ \forall y\colon \colon \ \mu_y \tag{40}$$

$$\lambda_y^i \epsilon \{0,1\} \forall y \tag{41}$$

Where Z_y^i represents a feasible solution of the investment capacity in year y, X_y^i are the values of the operational variables of year y that minimize the corresponding operational costs, with the corresponding value C_y^{opt}. The tuple $(Z_y^i, X_y^i, C_y^{opt})$ of known values is called a column and represents an extreme point of the RMP. Hence, the objective of the RMP is to determine the optimal combination of columns, λ_y^i, that represents a valid solution of the TNEP problem, i.e. where the yearly investment decisions Z_y^i are coherent with the cumulative capacities Z_y^i, enforced by (39). Equation (40) enforces that, for each year, only one column is selected. Note that, by solving the master problem a feasible solution of the original problem is obtained. Note that π_y and μ_y are the dual variables associated with constraints (39) and (40), respectively.

With this formulation of the RMP, within CG the subproblems, one for year y, have the following form:

$$z_y^{sp} = \min_{X,Z} C_y^{op} X_y - \pi_y^t Z_y \tag{42}$$

Subject to:

$$D_y X_y \leq Z_y \tag{43}$$

$$0 \leq X_y^i \leq X_y^{up} \tag{44}$$

Note that in the subproblems (42) - (44), the costs associated to the cumulated capacities is π_y^t. The CG algorithm solves TNEP iteratively, starting from a feasible solution (column). In each iteration, the RMP (39) - (41) is solved and the reduced costs π_y are computed for each year. Then, these reduced costs are utilized to formulate the subproblems. Since the dual variables π_y provide an estimation of the impact in the objective function (total costs of the system) of having additional cumulated capacity, using the dual variables π_y^t in the subproblems allows obtaining the largest improvement in the objective function in each iteration [52]. Another important characteristic of this formulation is that the subproblems can be solved individually for each year, which enables solving them in parallel.

The solution of each subproblem provides reduced costs (or slackness). If the reduced cost associated to a new column (i.e. feasible solution) is negative, then this column can improve the solution of the RMP and therefore it is added to the RMP. If none of the pricing sub-problems generate columns associated with a negative reduced cost, then the algorithm ends with an optimal solution of the original problem, since no column can further improve the current solution.

Since the RMP is a more restricted version of the original problem the value objective function in each iteration is an upper bound of the optimal solution. In other words, the solution of the RMP is a suboptimal feasible solution of the TNEP problem. On the other hand, the lower bound in each iteration can be obtained with equation:

$$Lb = \sum_{y=1}^{Y} z_y^{sp} - \mu_y \tag{45}$$

The algorithm can stop either when no new columns are added to the RMP in an iteration, or when the gap between the upper and lower bounds are within an acceptable margin.

2.3. Accelerating the performance of decomposition techniques

Despite the effectiveness of state-of-the-art decomposition techniques for solving large-scale optimization problems, the TNEP problem is still np-hard, i.e., there is no known algorithm that can find the optimal solution in polynomial time. This limits the scope of the case studies since simplifications in the models and the degree of freedom must be made. For example, in the TNEP problem, usual simplifications are using reduced network representations, limiting the planning horizon limiting the number of representative operational conditions and neglecting operational constraints. In response to this challenge, the development of novel algorithms that can find the optimal solution more efficiently is an active research area.

For CG, in [53] it was shown that adding multiple columns to the master problem in each iteration helps improving convergence time. In [54], to speed up the convergence, the authors reduced the complexity of the sub-problems by considering only promising integer variables, identified from the results of previous iterations. The values of non-promising variables are fixed to either zero or one. Even though this strategy reduces the time required to solve the subproblems, it may fail to converge if fixing a wrong value of the non-promising variables.

For Benders Decomposition, in [55] a method called maximum density cut (MDC) generation is proposed. The density of a cut is measured by the number of decision variables contained in a cut. The more decision variables contained in a cut, the larger portion of the feasible region is eliminated, which may help reducing the number of iterations required to converge to an optimal solution. The main limitation of this proposal is when the auxiliary problem that finds the MDC becomes to computationally expensive, to locate the MDC a MILP problem has to be solve, this can be done relatively quickly for certain optimization problems, however for larger scale problems like the ones found in transmission planning this could not be the case even if they have block diagonal structure. In [56], the authors propose using local branching to improve the efficiency of Bender's algorithm. Local branching consists of exploring the neighborhood of the solutions obtained from the master problem, in order to find new solutions. By doing so, multiple cuts are obtained in each iteration, thus increasing the chances of obtaining better upper bounds and generate optimal cuts that can strengthen or replace Bender's feasibility cuts. However, this approach requires solving difference instances of the master problem on each iteration, so it is only effective in optimization problems where the master problem does not become too complex with increasing iterations. Otherwise, solving the master problem could become a bottleneck for the algorithm's performance. Another strategy is presented in [57], where instead of generating cuts from the subproblems in the traditional way, so-called partially relaxed cuts are generated by solving a linear version of the master problem. Although partially relaxed cuts are less efficient (less tight) than normal cuts, they can be generated much faster. This can lead to a significant reduction in computation time for the early stages. Then, in later iterations, the relaxation is removed variable by variable, until the algorithm converges.

For Lagrangian relaxation, the usual approach is to update the dual multipliers using a subgradient-based approach, which can be complex and may lead to longer convergence times. To address this issue, in [58] the authors use particle swarm optimization (PSO), an evolutionary algorithm, to update the Lagrangian multipliers. This proposal was used to solve the scheduling problems in the industrial internet hybrid service flow scheduling. The problem with this approach is that the PSO, or any other evolutionary algorithm, must be properly tunned in order to be more efficient that a traditional gradient-based approach, which can be very time consuming. In addition, evolutionary algorithms work best when the subproblems can be solved with low computational effort, so that a large number of possible solutions can be explored. In [59], the authors propose simplifying the RPP for solving the p-median problem, a non-convex MILP problem. The p-median problem consists of allocating p facilities and assigning them to the demand points, such that each demand point is mapped to a single facility and that the sum of the weighted distance between all demand points and the corresponding facilities is minimized. The simplification consists of relaxing the equality constraints into inequalities, allowing for a broader range of solutions that still adhere to the problem's constraints but in a less rigid manner. As shown, this relaxed version of the RPP can improve the performance of the algorithm because the weaker enforcement of the equality constraints makes the inner minimization problem easier to solve and allows for the inclusion of a broader set of feasible solutions within the relaxation, potentially leading to a better approximation to the optimal solution of the original problem. Furthermore, this approach reduces the integrality gap for a large number of Lagrange multipliers, meaning that the semi-Lagrangian relaxation can achieve the same optimal solution as the original problem.

2.4. Machine learning applications on large scale optimization problems

In recent years, several researchers have started to use the capabilities of machine learning to solve large-scale complex optimization problems, these attempts can mainly be divided into 3 categories [60]: End-to-end learning, tailor-made solutions based on reinforcement learning and algorithm specific solutions. End-to-end learning uses the capabilities of machine learning to approximate functions. In this case, the underlying system is treated as a black box. In the case of neural network large amount of data are used to create a (non-linear) map between a reduced number of inputs to the optimal outputs. These methods can be very useful to solve linear optimization problems, where the optimality can be even guaranteed. Furthermore, promising developments have been seen for solving non-linear problems with minimal optimality gap [61]. However, utilizing an end-to-end strategy to solve mixed-integer optimization problems with global optimality guarantees remains a challenge. On the other hand, reinforcement learning consists of training an agent to make decisions based on previous experience. The training method aims at maximizing a reward signal for good decisions, based on the specific objective function of the problem. Examples of this approach can be found on [62] [63] [64] in the area of wireless communication. For the TNEP problem, reinforcement learning was used in [65]. In this work, a Double Deep Q Network (DDQN) is trained to judge the aggregated value that candidate lines for expansion add to the power system. This information is then utilized to select the most appropriate expansion projects. As a result, the authors show that the proposed method can be very efficient in solving both, static and multi-year large-scale TNEP problems. However, the drawback of this proposal is that the optimality of the solution cannot be guaranteed.

Finally, algorithm specific solutions take advantage of the conventional capabilities of machine learning, such as binary classification or linear regression, to assist or replicate known, globally optimal algorithms to achieve faster solutions with limited or no loss of optimality. An example of this strategy can be found in [66], where an imitation learning technique is utilized to train a binary classifier for proposing an auxiliary prune policy for the branch and bound algorithm. The classifier identifies branches that should and should not be explored. The input data of the agent are features of the branch and bound search tree, which can be generalized to be applied to any kind of optimization problem of MILP nature. The results of this work showed that the proposed algorithm reduced the computation time with only a small loss in the optimality gap. In [67], for a Benders decomposition algorithm, an agent was trained to identify between good and bad quality cuts. Only quality cuts are added to the master problem, which helps keeping the computational complexity of the master problem to a minimum and, consequently, to speed up the convergence with no loss of optimality. Note that a common bottleneck in Bender's decomposition is solving the master's problem with increasing iteration, given the large number of cuts added. Other works that combine machine learning techniques with column generation are [54] [68] [69]. Particularly, in [68] a machine learning-based model trained to predict the reduced cost of the columns, without solving the pricing sub-problems. Then, only those columns with the most negative reduced costs are calculated. Note that solving the sub-problems is usually the most time-consuming part of the CG algorithm. This method showed an improvement in computing time with a small reduction in optimality. However, a drawback of this method is that it uses a greedy search policy for selecting the new columns, focusing only on the ones with the most negative reduced cost, which could not always be the best criteria for selecting new columns as other alternatives could results in less iterations on the long run [70]. Researchers in [70] suggest that diversifying the columns can reduce the number of iterations. Another approach was presented in [71], where a model learns to select the best columns generated at each iteration from the conventional pricing sub-problems, in order to reduce the computing time spent to optimize the restricted master problem. However, this strategy may not be suitable for the TNEP problem, where the bottleneck lies in solving the subproblems and not the restricted master problem. All these works show the great potential of algorithm specific learning. The proposals can be easily applied to various kinds of optimization problems and, in many cases, achieve solutions with minimal loss of optimality. However, only the proposal in [67] for Bender's decomposition can guarantee global optimality.

3. Proposed methodology

The main objective of this thesis is to propose a novel optimization algorithm that combines the use of artificial intelligence techniques with Column Generation to solve the TNEP problem. The proposed algorithm must ensure the optimality of the solutions found and require less computational time to converge, compared to the traditional Column Generation algorithm. The reason for choosing Column Generation is because the master problem is computationally inexpensive compared with the subproblems, which are the main target of our proposal. And also, to take advantage of an existing tool for the TNEP problem, called, FlexTran, developed by the Department of Electrical Engineering of the Universidad Federico Santa María and the Universidad de Chile.

As presented in section 2.2, the bottleneck of CG is to solve the pricing subproblems, of MILP nature. As a results of these subproblems, new columns are generated, which are used for solving the RMP. The main requirement for the algorithm to converge is to have enough good quality columns. Hence, a sound strategy for addressing this challenge is to identify good quality columns without solving the original MILP subproblems, but a relaxed version of them, of LP nature, i.e. where the investment variables are treated as continuous. Then, relevant features can be determined from the results of the relaxed subproblems, which can be used to train an agent that decides on the integer nature of the investment decisions. Once the agent decides on the values of the investment variables, the subproblems can be solved again, but this time fixing the complicated investment variables, i.e., another LP problem is solved. If the accuracy of the agent in predicting the optimal value of the investment decisions is high, this strategy allows to generate good quality columns in significant less amount of time: only two LP must be solved for each subproblems, instead of one MILP. A significant advantage of this strategy is that the optimality of the solution can also be guaranteed. Indeed, a lower bound and upper bound is obtained on each iteration from the subproblems and master problem respectively. Therefore, the optimality gap can be measured at any point and furthermore in any iteration a new set of columns can be generated to improve the current gap.

An overview of the proposed methodology for improving the CG algorithm with machine learning is presented in Figure 3:

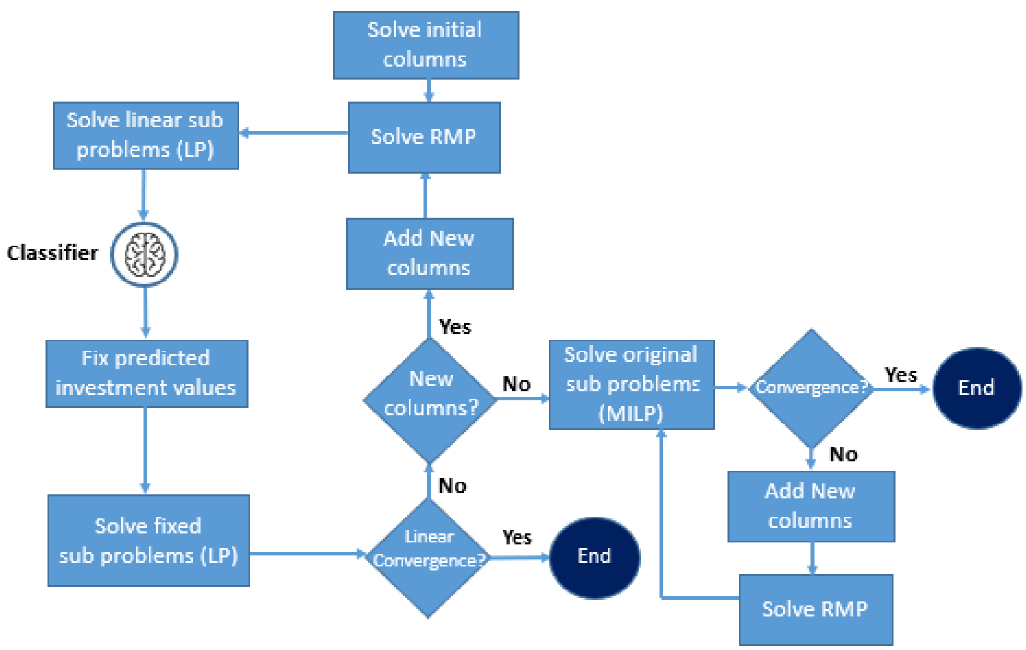

Figure 2: Interaction diagram of the binary classifier with the column generation algorithm.

As presented above, the main challenge of this proposal is to find good quality columns, from relevant features that allows the binary classified to predict the true value of the investment decisions from the results of the (relaxed) linear subproblems. In this thesis, the following features will be explored: 1) the usage ratio of each expansion candidate (lines, transformers and batteries), 2) the values of the dual variables associated with the minimum and maximum capacity of each expansion candidate, 3) the expected economic benefits of each expansion candidate and 4) the value (continuous) associated to the investment variables from the LP solution. Next, a brief description of each of these features are presented.

Usage ratio: As its name suggests, the usage ratio quantifies how much of the total capacity of the project is being used. A high usage ratio (for example, in a line) suggests a lack of capacity and, consequently, a need for expansion. On the other hand, a low usage ratio indicates that the capacity is not fully used and, therefore, there is little or no incentive for expansion.

Since the TNEP is solved for several representative days within a year, the average value is used. The TNEP problem addressed in this thesis contemplates two kinds of transmission investment projects: new lines or battery storage devices. However, for both kinds of projects the concept of usage ratio works in the same way. For transmission lines, to compute the usage ratio the absolute value of the power flows is used. For storage devices, the usage ratio is computed using the power charge (or discharge) in the hours where it is charging or discharging divided by the max power capacity of the device.

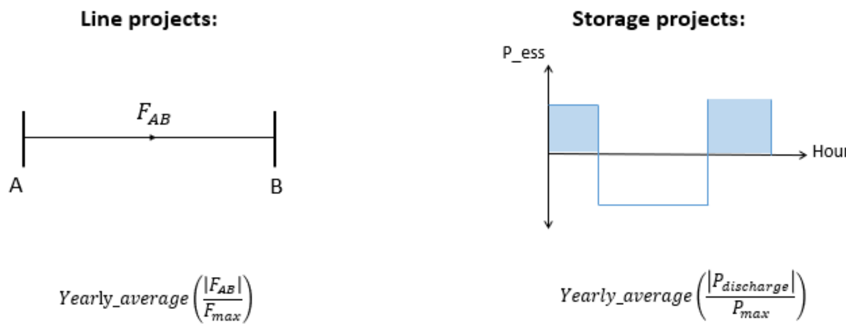

Figure 3: Usage ratio for transmission line projects (left) and for storage devices (right).

Positive and negative capacity limit duals: The dual values associated with the positive and negative capacity limits of the investment projects not only indicates of how long the project is operating at maximum capacity, but also it indicates the economic benefit of increasing said capacity. Hence, the higher this value, the stronger the indication for expanding this project. To consider several operating conditions within a year, the feature contains the sum of the dual values over all hours, and to compare the benefit with the investment cost of the project the value of π dual is used, resulting in the following formula:

$$Cap\_dual_{l,y} = \pi_{y,l} + \sum_{h=1}^{24}\sum_{d=1}^{D}\left\|\mu_{l,h,d,y}^{max} + \mu_{l,h,d,y}^{min}\right\| \cdot F_l^{max} \tag{46}$$

$$Cap\_dual_{ess,y} = \pi_{ess,l} + \sum_{h=1}^{24}\sum_{d=1}^{D}\left\|\mu_{ess,h,d,y}^{max} + \mu_{ess,h,d,y}^{min}\right\| \cdot P_{ess}^{max} \tag{47}$$

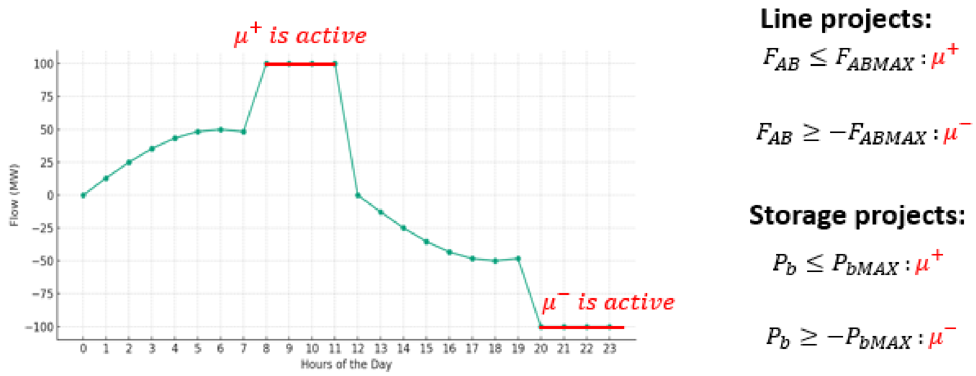

Figure 4: Positive and negative capacity limit dual values in the TNEP problem.

Figure 4 shows an example of the typical flow in power line with capacity limits in the positive and negative direction. When the flow on the line reaches the limit at some hour the associated constraints become binding and thus the capacity duals are different than zero (active). For every other hour when the flow of the line is below the limit the capacity duals are equal to zero (inactive).

Expected economic benefit: This feature estimates the economic benefits of investing in the project, as the difference between the expected incomes and the annualized investment costs. Note that, for CG, the annualized investment costs in each iteration, for each expansion project, is given by the dual variable π of constraint (39). For transmission line projects, the expected incomes associated with the project can be computed using the yearly congestion revenue from the locational marginal price difference at the extreme nodes. For storage projects, the expected income can be estimated by computing the value of the charging and discharging energy, from the marginal costs of the corresponding busbar. Similar to the previous feature, the expected economic benefits are normalized to ensure good performance of the classifier in different problems. The mathematical formulation for the expected economic benefits of lines (48) and storages (49) is the following:

$$Eb_{l,y} = \pi_{y,l} + \sum_{h=1}^{24} \sum_{d=1}^{D} \left\| (LMP_{h,d,y}^{A} - LMP_{h,d,y}^{B}) \cdot F_{h,d,y}^{AB} \right\| \tag{48}$$

$$Eb_{ess,y} = \pi_{ess,l} + \sum_{h=1}^{24} \sum_{d=1}^{D} P_{h,d,y}^{ess} \cdot LMP_{h,d,y}^{A} \tag{49}$$

Figure 5 presents an illustrative example how the features used by the binary classified to estimate the true value of the investment decisions, based on the results of the relaxed subproblems. Here we only show a projection into a 2-dimensional plane but since we have 4 features the real decision frontier will be in a 4-dimensional space. Where the machine learning agent will learn a proper hyperplane to separate the data points into the 2 classes.

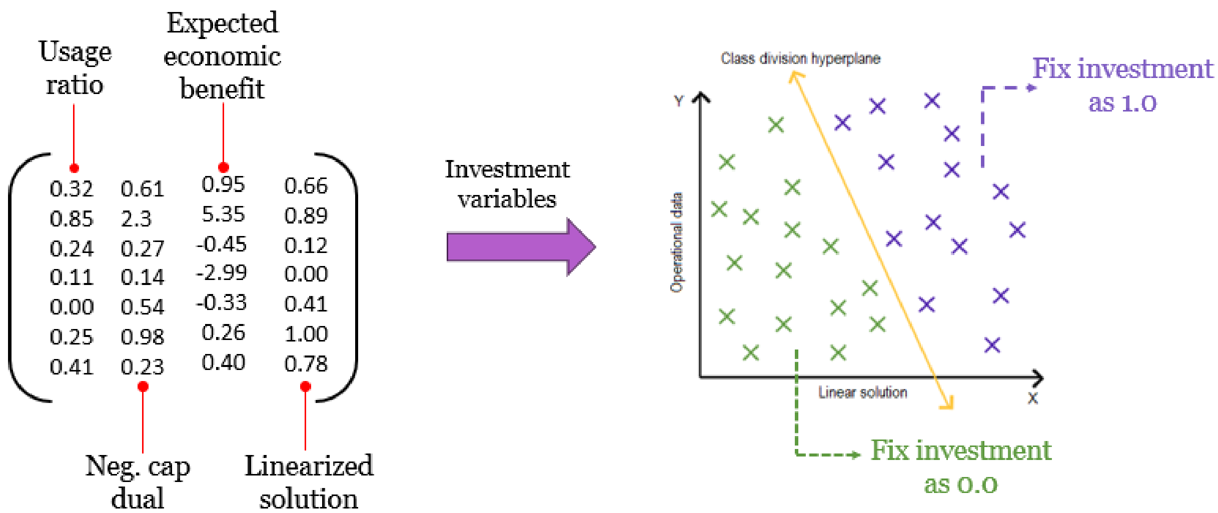

Figure 5: Example of class separation from features by a 2-dimensional classifier.

Binary classification is a fundamental task in machine learning where an agent learns to distinguish between 2 classes from different features in the dataset. There exists a variety of techniques for this kind of tasks, so the appropriate method must be selected for the specific problem addressed on this research. The following binary classifications methods will be considered in the development of this thesis:

Support Vector Machines (SVM): SVM is primarily a linear classifier that aims to find the best hyperplane separating the data into two classes. The "best" hyperplane is the one that maximizes the margin between the closest points of the two classes, termed support vectors. In cases where the data isn't linearly separable, SVM can still be applied using an appropriate kernel which is a nonlinear mathematical function that transforms the features, the kernel maps the data to a higher-dimensional space where it might be linearly separable. The main advantage of these techniques is its flexibility since a variety of kernels can be selected for different nonlinear problems, the most common alternatives are the radial basis kernels (rbf), the polynomial kernels and the sigmoid kernels. Meanwhile, the main disadvantage of this method is that it can be computationally expensive on large datasets.

Decision Trees: A decision tree is a flowchart-like tree structure where each internal node represents a feature (or attribute), each branch represents a decision rule, and each leaf node represents an outcome or class label. The paths from the root to the leaf is the learned classification rule. The focus of the decision tree is to minimize the entropy which is a measure of how even the data on a given node is and thus create the best possible separations of the data. In the decision space this can be seen as a frontier that splits sectors of the plane according to each rule. Because of this complex decision frontier decision trees archive high performance in a variety of linear and nonlinear classification problems. However, this makes decision trees more sensitive to small variations in data and prone to overfitting. To avoid these issues appropriate separation of training and testing data and balancing of classes is essential.

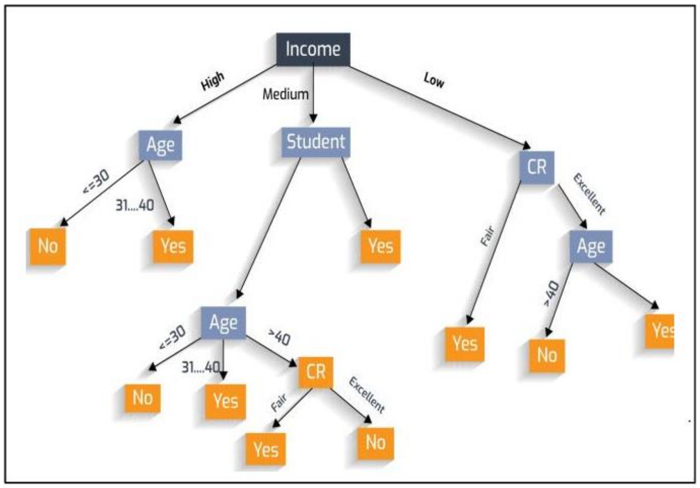

Figure 6: Example of a decision tree.

Figure 6 shows an example of how decision trees classify data by making sequential splits on the datapoints depending on the values of a specific features.

Random Forest: A Random Forest is an ensemble of decision trees, but instead of just averaging the output of those trees, it leverages two key concepts: bootstrapping and feature randomness. This combination ensures that each decision tree in the ensemble is trained on a slightly different dataset and based on different features, leading to a diverse set of trees, and thus reducing the chance of overfitting.

Bootstrap Aggregating (Bagging): For each tree in the ensemble, a dataset is created by randomly selecting samples from the original dataset with replacement. This means some samples may be repeated, while others might be left out. The idea of bagging is to avoid overfitting by diversifying the training data of individual trees.

Feature Randomness: When determining splits in each tree, only a random subset of features is considered, adding more diversity to the tree structures, and making the forest more robust.

Both methods help to address the main disadvantages of decision trees which are overfitting and data sensibility and thus making this technique more reliable for a variety of classification problems. Another advantage of this method is that the classification decision is defined by a voting system among the many random decision trees and thus a probability of the prediction can be obtained as a measure of how sure the agent is about the classification of a given point into a specific class.

Neural networks: In binary classification, a neural network (NN) aims to distinguish between two classes, often labeled as 0 and 1. For a NN the fundamental building block is the "neuron" or node. Each neuron processes an input, which can have multiple dimensions, then applies a nonlinear function to the weighted sum of its input and produces an output. The network's goal is to adjust its weights and biases during training to learn the underlying patterns in the data and make accurate classifications. A typical NN architecture consists of an input layer which receives the features of the data. Then a number of hidden layers which contains neurons that perform computations and transformations on the data. The depth (number of hidden layers) and width (number of neurons in each layer) can vary depending on the complexity of the task and the dimensions of the data. Finally for binary classification, the output layer typically has a single neuron. This neuron produces a value, typically between 0 and 1, representing the probability of the input belonging to class 1. An activation function, such as the sigmoid function, is used in this layer to squash the output into the (0,1) range.

For training the neural network a loss function must be used to adjust the weights and biases in proportion to their contribution to the measured error (loss). In binary classification, the most commonly used loss function is binary cross-entropy. It measures the difference between the actual labels and the predicted probabilities. The network's objective during training is to minimize this loss function. This means that the training process is a kind of optimization problem where the weights are updated according to the backpropagation rule which uses optimization techniques like gradient descent to ensure that the prediction error is reduced over subsequent iterations. The main disadvantage of neural network approach to binary classification is the lack of interpretation of the decision made by the agent in contrast with decision tree-based methods and the fact that they usually require larger volumes of data and computational resources to train effectible.

To evaluate the accuracy and the generalization ability of the proposal, two network models are used the HRP-38 busbar system [72] and the SEN transmission network. The agent is tested on a scenario blind strategy, meaning that the agent is trained using a specific set of scenarios for the expansion planning problem (for example, different combinations demand growth, renewable energy penetration, fuel costs, hydrology, etc.) and then the performance is evaluated using a different set of scenarios for the same network. Finally, the performance of the agent is measured as the average accuracy across all scenarios using cross validation. Figure 6 shows and example of how cross validation is applied in this proposal, in this example model is trained with the solution data from scenarios 2,3 and 4 and we test the prediction accuracy of the trained agent on scenario 1. This will give an average accuracy for this model. Then for model 2 the target scenario is switched to scenario 2 and the model is trained again we different data (scenarios 1,3 and 4) this will give another average accuracy value. Once all scenarios have been evaluated with this method the total average accuracy and variance are calculated from all scenarios.

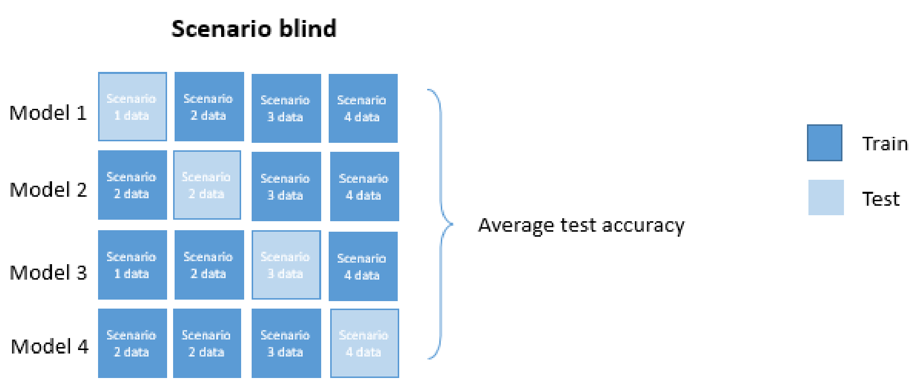

Figure 7:Scenario blind methodology for testing model accuracy.

The different scenarios are built from different combinations of the input parameters (number of operating conditions, planning horizons, hydrology, among other parameters). The following table presents a summary on how the scenarios are created and divided for training and testing on the small 38 bus study case:

Scenario	Storage Cost	Demand Growth %	Reneawble Capacity Growth %	Dataset
1	Low	2.5	10	Training
2	Low	4	10	Training
3	Low	1.2	10	Training
4	Mid	2.5	8	Training
5	Mid	4	8	Training
6	Mid	1.2	8	Training
7	High	2.5	5	Training
8	High	4	5	Training
9	High	1.2	5	Training
10	Low	4	8	Training
11	Mid	2.5	10	Training
12	High	4	8	Test
13	Low	1.2	5	Test
14	Mid	1.2	10	Test
15	High	2.5	10	Test
16	Low	4	5	Test
17	Mid	1.2	5	Test
18	High	4	10	Test
19	Low	1.2	8	Test
20	Mid	4	10	Test
21	High	2.5	8	Test

Figure 8: Summary of different scenario combinations.

4.Results

The agent is trained by solving different scenarios of the TNEP problem on a 38-bus test system (Appendix 1) considering a 15-year planning horizon and with 12 representative days. This system contains a mix of RES and thermal generators but not dam hydro power, over 100 existing transmission lines, 20-line projects and 10 storage projects. The results of the agent accuracy evaluated on the different scenarios of the 38-Bus system using a classifier without confidence threshold and another agent with a 90% confidence threshold are presented in the following tables:

Table 1: Accuracy results of the binary classifier

without all data points for the 38-bus system.

Scenario	Accuracy
1	93.72%
2	90.76%
3	91.90%
4	85.17%
5	90.68%
6	93.73%
7	86.72%
8	92.62%
9	89.66%
10	91.25%
Average	90.50% ± 1.68%

Table 2:Accuracy results of the binary classifier with datapoints

inside the 90% confidence threshold for the 38-bus system.

Scenario	Accuracy
1	98.91%
2	98.63%
3	97.18%
4	98.75%
5	99.21%
6	99.23%
7	97.46%
8	97.95%
9	97.60%
10	98.61%
Average	98.35% ± 0.44%

The results in Table 1 contain the accuracy of the agent over all data points, while Table **2** only contains results for data points within a 90% confidence threshold. Despite this threshold being particularly high, most data points fall inside this threshold, meaning that the results in Table 2 only contain 15% less data points compared to Table 1.

Then the algorithm was tested in 300 busbars system based on the Chilean national transmission network. The model is not only larger in the number of buses, lines, and generator but it also considers hydro network management and new expansion project alternatives like line voltage uprating and reconductoring. It will also consider a larger planning horizon of 20 years and the same 12 representative days. The agent is trained using solution data taken from solving different scenarios of this network model creating scenarios with different hydrological series, demand, and RES profiles. The test on this larger system rendered the following results:

Table 3: Accuracy results of the binary classifier

with 90% threshold for the 300-bus Chilan system.

Scenario	Accuracy
1	96.42%
2	88.07%
3	91.18%
4	97.08%
5	96.17%
6	96.13%
7	95.21%
8	87.89%
9	92.68%
10	96.11%
Average	93.69% ± 0.04%

Table 4:Accuracy results of the binary classifier with 90% threshold for the 300-bus Chilan system.

Scenario	Accuracy
1	97.84%
2	93.70%
3	94.11%
4	98.16%
5	97.49%
6	97.94%
7	96.37%
8	96.62%
9	93.98%
10	97.53%
Average	96.37% ± 0.02%

The results of table 3 and 4 show evidence of high accuracy in the predictions across multiple scenarios for the largest study case as well. Furthermore, the values presented for both study cases show an average accuracy that is above what would be achieved with only a linear regression from the LP solution of the investment variables. The accuracy achieved only by linear regression was found to be and 85.65% for the 38-bus system and 86.22% for the larger study case.

Even though the proposal has shown a high accuracy in predicting the investment variables from features of the LP problem, and an improvement over the linear regression of the LP solution. The value of the proposed algorithm still has to be tested by solving the TNEP problem with the classifier and proving that is capable of consistently achieving faster convergence into a globally optimal solution.

The computation time of the ML assisted algorithm with the pretrained binary classifier integrated into column generation was compared with the performance of the original CG algorithm. The following figure shows the convergence speed achieved by an agent predicting all data points without using a confidence threshold, and the results for an agent predicting only variables inside a 90% confidence threshold, solving for the remaining low confidence variables as integers (reduced MILP subproblems):

Table 5:Solution time summary on the 38-bus system without confidence threshold.

Without confidence threshold			
Scenario	CG algorithm [min]	ML algorithm [min]	time save %
1	63.48	32.55	48.72
2	60.49	43.43	28.20
3	55.67	41.88	24.77
4	51.96	56.41	-8.56
5	56.65	41.98	25.90
6	64.16	44.35	30.88
7	55.29	75.17	-35.96

8	73.99	42.95	41.95
9	55.84	51.19	8.33
10	71.99	42.63	40.78
Average	60.95 ± 7.03	47.25 ±11.02	20.50 ± 24.62

Table 6: Solution time summary on the 38-bus system with 90% confidence threshold.

With 90% confidence threshold			
Scenario	CG algorithm [min]	ML algorithm [min]	time save %
1	63.48	31.78	49.94
2	60.49	34.47	43.02
3	55.67	56.63	-1.72
4	51.96	41.41	20.32
5	56.65	42.69	24.64
6	64.16	61.55	4.07
7	55.29	56.71	2.57
8	73.99	43.53	41.17
9	55.84	43.43	41.24
10	71.99	34.04	52.72
Average	60.95 ±7.77	44.62 ±9.87	27.28 ± 20.32

The results reported for both scenarios are obtained while achieving global optimality, meaning that the CG algorithm ends when it can no longer generate any new columns with negative reduced cost and thus has closed the optimality gap because the upper bound is equal to the lower bound.

Lastly the same test was carried for the larger study case over 10 different scenarios. However, in this case given the size and complexity of the problem it was not possible to solve all the MILP subproblem with zero optimality gap in reasonable computation time. To avoid this issue, it was necessary to use a 4-hour time limit on the solver to ensure that a sufficiently good solution for the subproblems was achieved by the CG algorithm. This meant that in this case it was not possible to achieve computational time saves using confidence thresholds and solving partially relaxed MILP subproblems, since they would take up to the solver time limit to find a solution even if the subproblem has significantly less integer variables. However, it was still possible to achieve significant time saves with the fully LP subproblems given the high accuracy of the classifier without confidence thresholds. The comparation of solution times for the 300-bus system using the classifier agent are presented in the following Table:

Table 7: Solution time summary for the Chilean transmission system.

With confidence threshold			
Scenario	CG algorithm [hours]	ML algorithm [hours]	time save %
1	12.34	28.59	56.84

2	17.39	24.78	29.83
3	17.53	25.03	29.96
4	7.27	16.35	55.55
5	13.71	15.46	11.32
6	12.19	20.15	39.50
7	16.60	24.25	31.55
8	12.58	19.54	35.62
9	12.42	20.37	39.03
10	20.14	34.78	42.09
Average	14.22 ±3.71	22.93 ±5.83	37.13 ± 13.19

The better understand the behavior of the agent integrated into the column generation algorithm the evolution of the objective over each iteration is presented for the worst and best scenario of the Chilean case study:

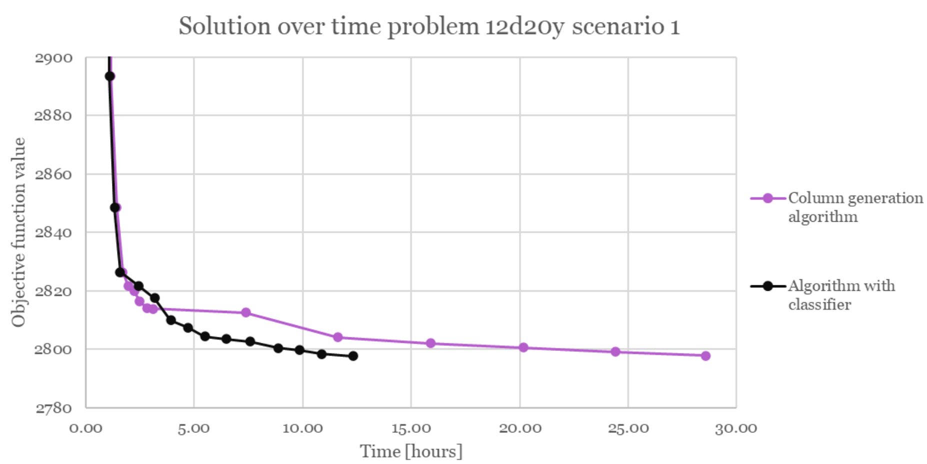

Figure 9: Evolution of the objective function in the best scenario for the Chilean transmission network.

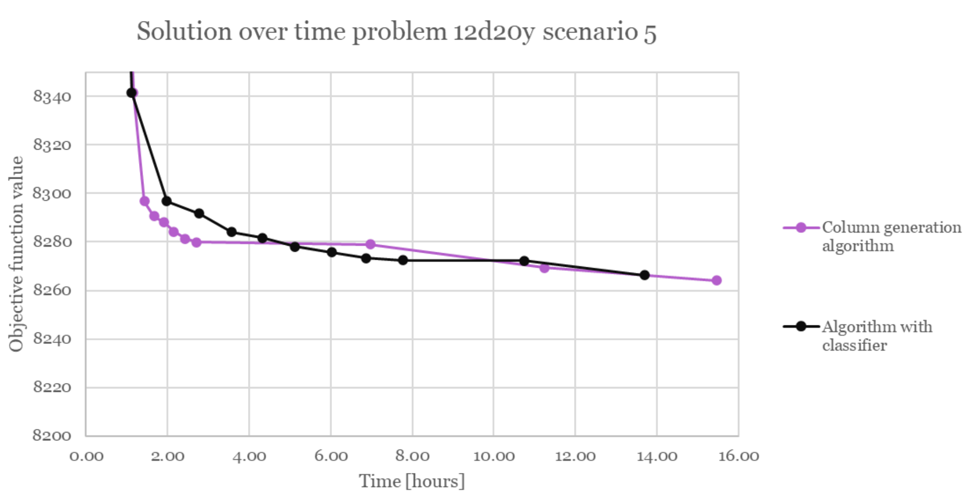

Figure 10: Evolution of the objective function in the worst scenario for the Chilean transmission network.

Figure 9 shows the results for the best scenario, meaning where we saw the biggest time saves, solving the problem 56.84% faster. Here we can see that when the agent is active, since we are solving only LP subproblems each iteration takes significantly longer. Furthermore, if the accuracy of the agent is high enough the quality of the columns being generated results in a significant improvement over the objective function and thus a stepper slope over time compared with the original column generation algorithm. The consequence of this is that the agent can achieve a solution of identical quality as the CG algorithm in less computation time. Meanwhile, in Figure 10 it can be seen that even though each iteration takes less time than the MILP iterations of the column generation algorithm, the overall slope of the objective function over time is comparable to the original algorithm since each iteration with the agent does not have a large enough improvement over the objective function. This is due to the lower accuracy of the classifier for the scenario that produces a new column of less quality. Even still the total computation time was still faster than the CG algorithm but not as significant es the case shown in Figure 9. Both these examples highlight the importance of the accuracy of the agent for this proposal to work as intended.

5.Discusion

We can see from Table 1 that the binary classification agent can predict the investment variables with over 90% accuracy in most scenarios, using all the variables across different scenarios of the TNEP problem. This indicates that the proposed features were relevant enough to inform the agent and enable accurate predictions. Furthermore, as shown in Table 2, the agent achieves even higher accuracy when focusing only on data points within a high confidence threshold. Here, we achieve an average accuracy above 98% and even improve the statistical error to a lower value of ±0.44%. This confirms that the data points within the high confidence threshold of the random forest binary classifier have higher prediction accuracy, as expected. Additionally, most of the data points are concentrated within this threshold (85% of data points), meaning that only a few integer variables must be solved in the reduced subproblems, while most of them will be fixed by the predictions of the agent. We can also observe that in both cases, the accuracy has low variability over the different scenarios, which could result in a robust agent that can be useful for different instances of the TNEP problem.

Examining Table 5, we can see that the binary classifier agent without a confidence threshold is indeed capable of achieving time savings over several scenarios, resulting in a time reduction of up to 48.72% in scenario 1. However, this agent still presents some scenarios where the computation time is higher than the regular CG algorithm, such as scenarios 4 and 7. These instances occur when the binary classifier is not capable of emulating the behavior of the original CG algorithm with enough accuracy, thus requiring several more iterations than necessary to close the optimality gap. Despite these cases, in most scenarios, the agent can achieve significant time savings by solving only LP optimization problems. However, the performance is not consistent since we see significant variation in the time savings.

On the other hand, Table 6 shows that the overall time savings are greater and more consistent for the agent using a 90% confidence threshold, achieving time savings of up to 52.7% in scenario 11. Furthermore, we can see that the worst-case scenarios, where the agent takes longer than the original CG algorithm, have significantly reduced the extra time difference. These results highlight the importance of a high-accuracy agent for this application and the relevance of an appropriate threshold between the confidence in the predictions and the number of binary variables remaining in the subproblems.

For the larger study case it can be seen from the results in Table 7 that the prosed algorithm can achieve considerable time saves for more complex instances of the TNEP problem. Where the classifier was capable of consistently reducing the computation time for solving the TNEP to global optimality, in fact the average time save was around 37.13% while the maximum time save register was 56.84%. And furthermore, there was no case in the tested scenarios where the algorithm with the classifier resulted in longer computation times.

6.Conclusions

In this work, we were able to design an algorithm that integrates a binary classifier machine learning agent into the CG algorithm for solving the TNEP problem, resulting in significant time savings without compromising global optimality. Our main hypothesis, that an agent could use relevant features from the solution of a linearly relaxed subproblems to learn and predict the solutions of a MILP problem with high accuracy, was demonstrated, achieving good results over several scenarios.

It was demonstrated the proposed algorithm is capable of generating a large volume of training data for the classifier with a limited number of scenarios. This is because one instance of the column generation algorithm contains several iterations, investment variables, and planning years, which result in a large amount of subproblem data.

Finally, this work showcases the importance of high accuracy predictions for this kind of application, observing significant variations in the results depending on the accuracy of the agent, and even encountering some cases where the algorithm resulted in overall worse performance. Additionally, we demonstrated that trading off accuracy with the number of relaxed variables is a viable strategy for reducing the complexity of the subproblems and resulting in more robust and significant time savings.

Overall, the proposed methodology is capable of achieving significant time savings compared with the CG algorithm over several scenarios without compromising the optimality of the solution thanks to the reduced time required for solving LP subproblems instead of MILP subproblems.

7.Future work

In future work, we will evaluate the performance of the proposed ML-assisted algorithm using additional features and generating more datapoints per iteration of the CG algorithm by randomly perturbating the π values on the subproblems. Both these methods combined can help to improve the accuracy of the agent and thus may result in a more significant performance boost.

Furthermore, one area that should be explored is the capacity of designing an appropriate model-blind agent. This means an agent that could achieve high enough prediction accuracy to produce time savings not only in different scenarios of the expansion problem but also in different network topologies (solving an entirely different network from the training data's original network). We suspect that this could be achieved by generating a greater volume of training data on different network topologies and designing a proper normalization technique for the features among different systems. Another interesting area of research would be to take advantage of the linear nature of the optimization problems produced by this proposal to implement some novel strategies that could reduce the time spent solving the linear subproblems and result in even greater time savings.

8. Acknowledgments

This work was partially supported by the Chilean National Agency for Research and Development, ANID/Fondap 1523A0006 (SERC-Chile), ANID/Fondecyt 1231739, ANID/Fondef ID20I10412, ANID/Fondef ID23I10277, and by the supercomputing infrastructure of the NLHPC (ECM-02). Also, we express our gratitude to the Cordinador Eléctrico Nacional, the Comisión Nacional de Energía (CNE) and the Agencia Nacional de Investigación y Desarrollo (ANID) for providing the funding for this work.

References

[1] Ministerio de energía de Chile, "Política enérgetica nacional," Santiago,Chile, 2022.

[2] C. Velásquez, D. Watts, H. Rudnick and C. Bustos, "A Framework for Transmission Expansion Planning: A Complex Problem Clouded by Uncertainty," *IEEE Power Energy Magazine,* vol. 14, p. 2029, 2016.

[3] G. K. Pall, A. J. Bridge, J. Gray and M. Skitmore, "Causes of Delay in Power Transmission Projects: An Empirical Study," *Energies,* vol. 13, no. 1, 2020.

[4] M. Meisam, S. Carlos, A. Majid and R. Rubén, "Transmission Expansion Planning: Literature Review and Classification," *IEEE Systems Journal,* vol. 13, pp. 3129-3140, 2019.

[5] L. Maulén, M. Castro, Á. Lorca and M. N. Pincetic, "Optimization-based expansion planning for power and hydrogen systems with feedback from a unit commitment model," *Applied Energy,* vol. 343, 2023.

[6] R. Araujo, S. P. Torres, J. P. Filho, C. A. Castro and D. V. Hertem, "Unified AC Transmission Expansion Planning Formulation incorporating VSC-MTDC, FACTS devices, and Reactive Power compensation," *Electric Power Systems Research,* vol. 2016, 2023.

[7] I. Konstantelos and G. Strbac, "Valuation of Flexible Transmission Investment Options Under Uncertainty," *IEEE Transactions on Power Systems,* vol. 30, no. 2, pp. 1047 - 1055, 2015.

[8] S. M. Rashid, E. Akbari, F. Khalafian, M. H. Atazadegan, S. Shahmoradi and A. Z. G. Seyyedi, "Robust Allocation of FACTS Devices in Coordinated Transmission and Generation Expansion Planning considering Renewable Resources and Demand Response Programs," *International Transactions on Electrical Energy Systems,* vol. 2022, 2022.

[9] H. Oh, "Optimal Planning to Include Storage Devices in Power Systems," *IEEE Transactions on Power Systems,* vol. 26, pp. 1118-1128, 2011.

[10] F. Valencia, R. P. Behnke, D. O. Villalba, C. Rahmann and R. Cifuentes, "Special Protection Systems: Challenges in the Chilean Market in the Face of the Massive Integration of Solar Energy," *IEEE Transactions on Power Delivery,* 2016.

[11] J. Porst, J. Richter, G. Mehlmann and M. Luther, "Operation of Grid Boosters in Highly Loaded Transmission Grids," in *IEEE International Conference on Power Systems Technology (POWERCON)*, Kuala Lumpur, Malaysia, 2022.

[12] M. P. González and M. A. Ríos, "Generation and Transmission Planning using HVDC Grids," in *IEEE*

PES Transmission & Distribution Conference and Exhibition, Lima, Peru, 2018.

[13] R. Alvarez, C. Rahmann, N. Cifuentes and R. Palma-Behnke, "Multi-Year Stochastic Transmission Network Expansion Planning Considering Line Uprating," *IEEE Access,* vol. 9, pp. 33075-33090, 2021.

[14] M. Moradi-Sepahvand and T. Amraee, "Hybrid AC/DC Transmission Expansion Planning Considering HVAC to HVDC Conversion Under Renewable Penetration," *IEEE Transactions on Power Systems,* vol. 36, pp. 579-591, 2021.

[15] G. Diaz, A. Inzunza and R. Moreno, "The importance of time resolution, operational flexibility and risk aversion in quantifying the value of energy storage in long-term Energy Planning Studies," *Renewable and Sustainable Energy Reviews,* vol. 112, p. 797–812, 2019.

[16] D. HAYSSAM, A. RAWAN, A. HIBATALLAH, B. SARAH, A. ALAA, F. ALICE, S. RAHAF, A. REEM and S. ABDULHAMIT, "An overview of machine learning-based techniques for solving optimization problems in communications and Signal Processing," *IEEE Access,* vol. 9, p. 74908–74938, 2021.

[17] A. Tawhid, T. Teotia and H. Elmiligi, "Machine learning for optimizing healthcare resources," *Machine Learning, Big Data, and IoT for Medical Informatics,* 2021.

[18] N. Nazareth and Y. V. R. Reddy, "Financial applications of machine learning: A literature review," *Expert Systems with Applications,* vol. 219, 2023.

[19] K. Tsolaki, T. Vafeiadis, A. Nizamis, D. Ioannidis and D. Tzovaras, "Utilizing machine learning on freight transportation and logistics applications: A review," *ICT Express,* vol. 9, 2023.

[20] A. Entezari, A. Aslani, R. Zahedi and Y. Noorollahi, "Artificial intelligence and machine learning in energy systems: A bibliographic perspective," *Energy Strategy Reviews,* vol. 45, 2023.

[21] C. Morariu, O. Morariu, S. Răileanu and T. Borangiu, "Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems," *Computers in Industry,* vol. 120, 2020.

[22] S. R. Khuntia, B. W. Tuinema, J. L. Rueda and M. Meijden, "Time-horizons in the planning and operation," *IET Generation, Transmission & Distribution,* vol. 10, no. 4, 2015.

[23] P. David, S. Enzo and J. Contreras, "A Three-Level Static MILP Model for Generation and Transmission Expansion Planning," *EEE Trans. Power Syst,* vol. 28, p. 202–210, 2013.

[24] Z. Fang, H. Zechun and S. Yonghua, "Mixed-integer linear model for transmission expansion planning with line losses and energy storage systems," *IET Gener., Transm. Distrib.,* vol. 7, p. 919–928, 2013.

[25] S. H. M. D. R. H., R. Shrestha and O. Fujiwara, "A mixed integer linear programming model for transmission expansion planning with generation location selection," *Int. J. Elect. Power Energy Syst.,* vol. 23, p. 285–293, 2001.

[26] J. Contreras and F. Wu, "A kernel-oriented algorithm for transmission expansion planning," *IEEE Trans. Power Syst.,* vol. 15, p. 1434– 1440, 2000.

[27] M. Meisam, M. Hassan and R. Rubén, "Reliability Effects of Maintenance on TNEP Considering Preventive and Corrective Repairs," *IEEE Trans. Power Syst.,* vol. 32, p. 3768–3781, 2017.

[28] N. Saeed, A. Alireza and Abdollahi, "Modeling Probabilistic Transmission Expansion Planning in the Presence of Plug-in Electric Vehicles Uncertainty by Multi-State Markov Model," *IET Gener., Transm. Distrib.,* vol. 11, p. 1716–1725, 2017.

[29] A. Behnam, D. Shahab, A. Nima, J. Shahram and K. Ahad, "Robust transmission system expansion considering planning uncertainties," *IET Gener., Transm. Distrib.,* vol. 7, p. 1318–1331, 2013.

[30] R. A. Jabr, "Robust transmission network expansion planning with uncertain renewable generation and loads," *IEEE Trans. Power Syst.,* vol. 28, p. 4558–4567, 2013.

[31] K. Kim, Y. Park and K. Lee, "Optimal long term transmission expansion planning based on maximum principle," *IEEE Transactions on Power Systems,* vol. 3, pp. 1494 - 1501, 1988.

[32] R. Romero and A. Monticelli, "A hierarchical decomposition approach for transmission network expansion planning," *IEEE Transactions on Power Systems ,* vol. 9, pp. 373 - 380, 1994.

[33] L. L. Garver, "Transmission Network Estimation Using Linear Programming," *IEEE Transactions on Power Apparatus and Systems,* vol. 7, pp. 1688 - 1697, 1970.

[34] M. Rider, A. Garcia and R. Romero, "Power system transmission network expansion planning using AC model," *IET Gener.Transm.Distrib.,* vol. 1, p. 731–742, 2007.

[35] C. W, F. B, M. T, B. J, B. G and Y. D, "Variable Renewable Energy in Long-Term Planning Models: A Multi-Model Perspective," A multi model perspective tech report National Renewable Energy Laboratory, Denver, Colorado, 2017.

[36] A. Pina, C. Silva and P. Ferrão, "Modeling hourly electricity dynamics for policy making in long-term scenarios," *Energy Policy,* vol. 39, pp. 4692-4702, 2011.

[37] M. Rahmani, G. Vinasco, M. J. Rider, R. Romero and P. M. Pardalos, "Multistage transmission expansion planning considering fixed series compensation allocation," in *IEEE PES General Meeting*, National Harbor, MD, USA, 2014.

[38] Z. Luburić, H. Pandžić and M. Carrión, "Transmission Expansion Planning Model Considering Battery Energy Storage, TCSC and Lines Using AC OPF," *IEEE Access ,* vol. 8, pp. 203429 - 203439, 2020.

[39] M. Esmaili, M. Ghamsari-Yazdel, N. Amjady, C. Y. Chung and A. Conejo, "Transmission Expansion Planning Including TCSCs and SFCLs: A MINLP Approach," *IEEE Transactions on Power Systems ,* vol. 35, no. 6, pp. 4396 - 4407, 2020 .

[40] T. Qiu, B. Xu, Y. Wang, Y. Dvorkin and D. S. Kirschen, "Stochastic Multistage Coplanning of Transmission Expansion and Energy Storage," *IEEE Transactions on Power Systems,* vol. 32, no. 1, pp. 643-651, 2016.

[41] P. A. Sánchez-Pérez, S. Kurtz, N. Gonzalez, M. Staadecker and P. Hidalgo-Gonzalez, "Effect of Time Resolution on Capacity Expansion Modeling to Quantify Value of Long-Duration Energy Storage," in *IEEE Electrical Energy Storage Application and Technologies Conference (EESAT)*, Austin Texas, USA, 2022.

[42] G. Diaz, A. Inzunza and R. Moreno, "The importance of time resolution, operational flexibility and risk aversion in quantifying the value of energy storage in long-term Energy Planning Studies," *Renewable and Sustainable Energy Reviews,* vol. 112, p. 797–812, 2019.

[43] H. Kim, S. Lee, S. Han, W. Kim, K. Ok and S. Cho, "Integrated Generation and Transmission Expansion Planning Using Generalized Bender's Decomposition Method," in *IEEE International Conference on Computational Intelligence & Communication Technology*, Ghaziabad, India, 2015.

[44] S. Dehghan, H. Saboori, A. Kazemi and S. Jadid, "Transmission network expansion planning using a DEA-based benders decomposition," in *18th Iranian Conference on Electrical Engineering*, Isfahan, Iran, 2010.

[45] V. Asgharian and M. Abdelaziz, "Co-Optimization of Generation Capacity Planning and Carbon Capture and Storage Using Benders Decomposition," in *IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, 2019.

[46] R. Mínguez, R. García-Bertrand, J. M. Arroyo and N. Alguacil, "On the Solution of Large-Scale Robust Transmission Network Expansion Planning Under Uncertain Demand and Generation Capacity," *IEEE Transactions on Power Systems,* vol. 33, pp. 1242-1251, 2018.

[47] A. Moreira, A. Street and J. M. Arroyo, "An adjustable robust optimization approach for contingency-constrained transmission expansion planning," in *IEEE Power & Energy Society General Meeting*, Denver, CO, USA, 2015.

[48] J. C. Villumsen and A. Philpott, "Column Generation for Transmission Switching of Electricity Networks with Unit Commitment," in *International MultiConference of Engineers and Computer Scientists*, Hong Kong, 2011.

[49] H. Chen, X. Wang and X. Zhao, "Generation planning using Lagrangian relaxation and probabilistic production simulation," *International Journal of Electrical Power & Energy Systems,* vol. 26, no. 8, pp. 597-605, 2004.

[50] J. Kim and Y. Kim, "A Lagrangian Relaxation Approach to Multi-Period Inventory/Distribution

Planning," *The Journal of the Operational Research Society,* vol. 513, pp. 364-370, 2000.

[51] J. A. Muckstadt and S. A. Koenig, "An Application of Lagrangian Relaxation to Scheduling in Power-Generation Systems," *Operations Research,* vol. 25, no. 3, pp. 387-403, 1977.

[52] A. J. Conejo, E. Castillo, R. Mínguez and R. García-Bertrand, Decomposition Techniques, 2006, pp. 112-113.

[53] N. Touati, L. L´etocart and A. Nagih, "accelerating convergence of column generation Technical report," University Paris, 2006.

[54] M. Morabit, G. Desaulniers and A. Lodi, "Machine-learning–based column selection for column generation," *Transportation Science,* vol. 55, p. 815–831, 2021.

[55] S. Georgios and I. Marianthi, "Speed-up Benders decomposition using maximum density cut (MDC) generation," *Annals of Operations Research,* vol. 210, p. 101–123, 2013.

[56] W. Rei, J.-F. Cordeau, M. Gendreau and P. Soriano, "Accelerating Benders Decomposition by Local Branching," *INFORMS Journal on Computing,* vol. 21, 2008.

[57] G. Vojvodic, L. J. Novoa and A. I. Jarrah, "Experimentation with Benders decomposition for solving the two-timescale stochastic generation capacity expansion problem," *EURO Journal on Computational Optimization,* vol. 11, 2023.

[58] Y. Song, W. Luo, P. Xu, J. Wei and X. Qi, "An improved Lagrangian relaxation algorithm based SDN framework for industrial internet hybrid service flow schedulling," *Scientific Reports,* vol. 12, 2022.

[59] C. T. P. V. C. Beltran, "Solving the p-Median Problem with a Semi-Lagrangian Relaxation," *Computational Optimization and Applications ,* vol. 35, p. 239–260, 2006.

[60] H. Dahrouj, R. Alghamdi, H. Alwazani, S. Bahanshal, A. A. Ahmad, A. Faisal, R. Shalabi, A. Subasi, O. A. Kittaneh and J. S. Shamma, "An overview of machine learning-based techniques for solving optimization problems in communications and Signal Processing," *IEEE Access,* vol. 9, p. 74908–74938, 2021.

[61] F. C. Beylunioglu, M. Pirnia, P. R. Duimering and V. Ganesh, "Robust Training for AC-OPF," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[62] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu and N. D. Sidiropoulos, "Learning to Optimize: Training Deep Neural Networks for Interference Management," *IEEE Trans. Signal Process,* vol. 66, p. 5438–5453, 2018.

[63] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," *IEEE J. Sel. Areas Commun,* vol. 37, p. 2239–2250, 2019.

[64] Z. Z. Xianfu Chen, C. Wu, M. Bennis, H. Liu, Y. Ji and H. Zhang, "Multi-tenant cross-slice resource orchestration: A deepreinforcement learning approach," *IEEE J. Sel. Areas Commun,* vol. 37, p. 2377–2392, 2019.

[65] Y. Wang, X. Zhou, H. Zhou, L. Chen, Z. Zheng, Q. Zeng, S. Cai and Q. Wang, "Transmission Network Dynamic Planning based on a double deep-Q network with Deep Resnet," *IEEE Access,* vol. 9, p. 76921–76937, 2021.

[66] M. Lee, G. Yu and G. Y. Li, "Learning to branch: Accelerating resource allocation in wireless networks," *IEEE Transactions on Vehicular Technology,* vol. 69, p. 958–970, 2020.

[67] M. Lee, N. Ma, G. Yu and H. Dai, "Accelerating generalized benders decomposition for wireless resource allocation," *IEEE Transactions on Wireless Communications,* vol. 20, p. 1233–1247, 2021.

[68] Y. Shen, Y. Sun, X. Li, A. Eberhard and A. Ernst, "Enhancing column generation by a machine-learning-based pricing heuristic for graph coloring," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 36, p. 9926–9934, 2022.

[69] N. Furian, M. O'Sullivan, C. Walker and E. Çela, "A machine learning-based branch and price algorithm for a sampled vehicle routing problem," *OR Spectrum,* vol. 43, p. 693–732, 2021.

[70] N. Touati-Moungla, L. Létocart and A. Nagih, "Solutions diversification in a column generation scheme," *Algorithmic Operations Research,* vol. 5, 2010.

[71] M. Morabit, G. Desaulniers and A. Lodi, "Machine-learning–based column selection for column generation," *Transportation Science,* vol. 55, p. 815–831, 2021.

[72] Z. Zhuo, N. Zhang, J. Yang, C. Kang, C. Smith, M. J. O'Malley and B. Kroposki, "Transmission Expansion Planning Test System for AC/DC Hybrid Grid With High Variable Renewable Energy Penetration," *IEEE Transactions on Power Systems ,* vol. 35, no. 4, pp. 2597 - 2608, 2020.

[73] A. Abeer, A. Atinuke, N. J., N. Ahlmahz and K. Daniel, "Selecting and evaluating representative days for generation expansion planning," *Power Systems Computation Conference (PSCC),* p. 1–7, 2018.

[74] P. Pena, N. Morales, M. Artenstein, A. Pizzini and C. Zoppolo, "Planning in transmission systems with a great level of penetration of distributed generation," *IEEE PES Innovative ,* 2015.

[75] G. A. Blanco, F. G. Olsina, O. A. Ojeda and F. F. Garces, "Transmission expansion planning under uncertainty Ȗ The role of FACTS in providing strategic flexibility," in *IEEE Bucharest PowerTech*, Bucharest, Romania, 2009.

[76] A. Wang, M. Xu, H. Zhu, T. Wang, Y. Wang, C. Meng, S. Liu, Y. Zheng and Y. Hu, "Study on Transmission Planning of Combined Wind and Storage System," in *2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing, China, 2018.

[77] S. Dehghan and N. Amjady, "Robust Transmission and Energy Storage Expansion Planning in Wind Farm-Integrated Power Systems Considering Transmission Switching," *IEEE Transactions on Sustainable Energy ,* vol. 7, pp. 765 - 774, 2016.

[78] T. Ding, C. Li, X. Liu, H. Xie, Y. Tang and C. Huang, "Optimally Allocating Energy Storage for Active Distribution Networks to Reduce the Risk Under N-1 Contingencies," *IEEE Systems Journal,* vol. 15, pp. 1518-1527, 2021.

[79] M. Rodrigo, S. Alexandre, M. Jos and P. Mancarella, "Planning Low-Carbon Electricity Systems under Uncertainty Considering Operational Flexibility and Smart Grid Technologies," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 375, 2017.

[80] D. H. Huanca and L. A. G. Pareja, "Chu and Beasley Genetic Algorithm to Solve the Transmission Network Expansion Planning Problem Considering Active Power Losses," *IEEE Latin America Transactions,* vol. 19, pp. 1967-1975, 2021.

[81] F. Liu, Y. Su, J. Gou, C. Li, W. Xu and Y. Liu, "A Bi-level transmission expansion planning model considering the electricity market," in *3rd International Conference on Electrical Engineering and Control Technologies (CEECT)*, Macau, Macao, 2021.