MDPI

Article

Predicting the Climate Change in different climate zones using Grammatical Evolution

Constantina Kopitsa[1], Ioannis G. Tsoulos[2,*], Vasileios Charilogis[3]

[1] Department of Informatics and Telecommunications. University of Ioannina, Greece k.kopitsa@uoi.gr
[2] Department of Informatics and Telecommunications. University of Ioannina, Greece itsoulos@uoi.gr
[3] Department of Informatics and Telecommunications. University of Ioannina, Greece v.charilog@uoi.gr
* Correspondence: itsoulos@uoi.gr

Abstract

Climate change is no longer a future or hypothetical issue, but a present reality with tangible consequences and impacts on the well-being not only of the planet itself but also of all living organisms that constitute the ecosystem. Since 1969, the global surface temperature has been surpassing its own record almost every year, with NASA's final measurement for 2024 indicating an increase of approximately 2.65° Fahrenheit or 1.47° Celsius, in global temperature. These measurements indicate that we have reached the thresholds established by the United Nations, scientists, and member states through the Paris Agreement of 2015. In this study, we seek to gain a comprehensive understanding of climate change, and more specifically of its environmental impacts across seven distinct regions located in different climatic zones. Particular emphasis is placed on highlighting the rate at which climate change has altered annual temperature patterns during the period 1995–2025. To achieve this, we evaluated multiple modeling techniques and proposed the use of Grammatical Evolution with rule-based construction as the most suitable approach. GENCLASS, demonstrates robust and systematic superiority over the considered baselines (MLP, RBF, SVM, and NNC) across all datasets, achieving the lowest average classification error and indicating strong cross-regional generalization. In our research, we will examine a broad spectrum of climatic zones, specifically focusing on the Congo region, Saudi Arabia, Greece, Turkey, Reykjavik in Iceland, the Arctic Svalbard archipelago, and Cape Town in South Africa. Our experiments yielded highly robust results, with an average error rate of 1.66% and a classification accuracy of 98.34% for predicting climate change categories. These outcomes were further confirmed through established validation techniques, reinforcing the reliability of the model's performance.

Keywords: Genetic algorithms; Grammatical Evolution; Machine learning; Climate change

Received:
Revised:
Accepted:
Published:

Citation: Kopitsa, C.; Tsoulos, I.G.; Charilogis, V. . Predicting the Climate Change in different climate zones using Grammatical Evolution. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

1. Introduction

The power of nature and its capacity to respond when ecological balances are disrupted has long been recognized. A striking early example appears in The Persians by Aeschylus, a classical tragedy written in 472 BCE that recounts the historical events of 480 BCE [1] [2]. The play depicts the environmental and human catastrophes that afflicted the Persian army during its chaotic retreat, following acts that violated the natural world. Historical sources describe these transgressions, including the "chaining" of the Bosporus through the construction of a floating bridge, the plundering of agricultural lands to sustain their vast forces, and the burning and devastation of forests and plains—actions intended to

deprive local populations of essential resources. Such accounts illustrate that every action elicits a reaction, even from nature itself, which does not remain a passive observer but acts as a regulating force that safeguards the planet's equilibrium.

Recent scientific consensus confirms that anthropogenic factors, especially fossil fuel combustion, deforestation, and industrial emissions,play a pivotal role in driving climate change and its associated extreme events.In summary, climate change has been linked to extreme natural disaster events, including heat waves and cold spells, droughts, tropical storms and high winds, river floods, coastal floods [6] and wildfires [7] [8]. In particular, according to the United Nations, climate change refers to long-term alterations in temperature and weather patterns. While such variations may occur naturally—stemming from fluctuations in solar activity or major volcanic eruptions—human activities have emerged as the predominant cause since the 19th century. This anthropogenic influence is primarily attributed to the extensive combustion of fossil fuels, including coal, oil, and natural gas [3]. As will be demonstrated through the data presented below, climate change has entered an irreversible trajectory, with each passing year bearing witness to phenomena that repeatedly set new temperature records. Nevertheless, notable efforts have been made in the past—through treaties, regulations, targets, and constraints—aiming to reduce carbon dioxide emissions in an attempt to intercept climate change at a critical threshold. Specifically, the 1992 UN Conference of Environment and Development set as its ultimate objective "the stabilization of greenhouse gas concentrations in the atmosphere" [4]. In 1997, the Kyoto Protocol was adopted with the aim of committing industrialized nations and economies in transition to limit and reduce greenhouse gas (GHG) emissions, based on individually agreed targets. The overarching Convention, however, merely requires these countries to implement mitigation policies and measures, and to submit periodic reports [5]. In 2015, the Paris Agreement articulated long-term goals intended to guide 195 Parties in substantially reducing global greenhouse gas emissions. Its central aim is to constrain the rise in global temperatures to well below 2°C above pre-industrial levels, while encouraging efforts to limit the increase to 1.5°C. Achieving these targets is recognized as a critical step towards significantly mitigating the risks and adverse impacts associated with climate change [9]. Although we briefly observed the emergence of more environmentally conscious actions in the past, today, in the year 2025, NASA's data indicate that the debate has shifted to whether humanity will surpass the critical threshold of 1.5°C or 2.7°F global temperature [10].

In the field of climate change projections, numerous studies have been conducted. The first scientifically documented prediction of climate change was made in 1896, by Svante Arrhenius, in his work "On the Influence of Carbon Acid in the Air upon the Temperature of the Ground" [11].The aforementioned scholar estimated that the combustion of fossil fuels could ultimately emit sufficient carbon dioxide to raise the Earth's temperature by approximately 5 - 6°C or 9 - 11°F. Since then, approximately 130 years have passed; nevertheless, the fundamental physics underlying these calculations has remained unchanged. What has evolved are the techniques employed in our projections, which have become considerably more detailed and precise [12]. It is evident that climate change is intrinsically linked to global warming and the increasing occurrence of natural disasters, such as wildfires, floods, droughts, extreme weather events, landslides, and storms. In this context, we will refer to several studies on climate change and forest fires, such as: Koulelis et al., in 2023, conducted a study reviewing the impacts of climate change on Greek forests, examining factors such as climate trends, forest management practices, biodiversity, genetics, insect activity, and wildfires, drawing upon data from the Scopus and Mendeley databases as well as official reports [13]. The findings of Rovithakis et al., in 2022, indicate that fire danger is projected to progressively intensify in the future, particularly under the high-end

climate change scenario [14]. Giannakopoulos et al., in 2011, reported that shifting climate conditions, marked by an average increase in minimum temperatures of approximately 1.3°C and a 15% reduction in winter precipitation—indicate an intensified risk of forest fires in the future [15]. Comparable studies have linked floods to climate change, such as: Knox, in 2000, reported that flood chronologies from multiple regions indicate that periods of rapid climate change tend to coincide with more frequent occurrences of large and extreme flooding events [16]. Milly et al., in 2002, found that the frequency of major floods increased significantly throughout the twentieth century [17]. Bronstert, in 2003, highlighted that anthropogenic activities contributing to heightened flood risk encompass river regulation practices, intensified land use and forestry, as well as greenhouse gas emissions that drive changes in the global climate [18]. Subsequently, we will discuss studies addressing drought and its linkage to climate change, including: Cook et al., in 2018, projected that rising temperatures will heighten both the risk and severity of drought across large parts of the subtropics and mid-latitudes in both hemispheres, as a result of regional declines in precipitation combined with widespread warming [19]. Mukherjee et al., in 2018, noted that while droughts are natural phenomena, climate change has generally accelerated hydrological processes, causing them to develop more rapidly and with greater intensity [20]. Yuan et al., in 2023, observed that flash droughts—developing unusually rapidly in contrast to the more gradual archetypal droughts, are increasingly becoming the new normal [21].

Consequently, research on rising temperatures has attracted significant scientific interest. Subsequently, we will present studies that focus on the field of climate change, specifically on the projection of such meteorological phenomena. Kretschmer et al. published in 2017 “Early Prediction of Extreme Stratospheric Polar Vortex States Based on Causal Precursors” [22]. Nowack et al., in 2018, authored “Using Machine Learning to Build Temperature-Based Ozone Parameterizations for Climate Sensitivity Simulations” [23]. Huntingford et al., in 2019, highlighted that artificial intelligence (AI) can build upon identified climate linkages to deliver improved warnings of impending weather phenomena, including extreme events [24]. Mansfield et al., in 2020, introduced a machine learning approach that leverages a unique dataset of existing climate model simulations to identify relationships between short-term and long-term temperature responses under different climate forcing scenarios [25]. Haggag et al., in 2021, demonstrated the application of the developed model by linking flood disaster data from the Canadian Disaster Database with climate change indices for Ontario, which were then used to train, test, and validate the model [26]. Haq, in 2022, developed and optimized the Climate Deep Long Short-Term Memory (CDLSTM) model to forecast temperature and rainfall values across all Himalayan states [27]. Slater et al., in 2023, reviewed recent advances in hybrid hydroclimatic forecasting and delineated key challenges and opportunities for future research [28]. Singh Bist et al., in 2024, research introduced an innovative method for predicting climate change by employing deep learning techniques [29]. Gautam et al., in 2025, advanced the examination of climate time series data and enhance its analysis through deep learning methods [30].

A method that constructs classification rules using the technique of Grammatical Evolution [31] is proposed here for the prediction of climate changes. The Grammatical Evolution procedure can be considered as a genetic algorithm [32], where the chromosomes are series of integer values the representing the production rules of the provided Backus-Naur form (BNF) grammar [33]. The Grammatical Evolution procedure has been used successfully in a series of practical problems, such as data fitting problems [34,35], issues appeared in economic problems [36], issues regarding network security [37], problems regarding the quality of water [38], medicine problems [39], evolutionary computation [40], problems that appear in data centers [41], solution of trigonometric problems [42], automatic

composition of music [43], construction of the architecture of neural networks [44,45], production of numerical constants [46], playing video games [47,48], energy problems [49], combinatorial optimization [50], cryptography problems [51], construction of decision trees [52], electronic problems [53] etc.

This paper proposes the application of a technique for automatically generating classification rules to predict climate change in various regions of our planet. These rules are expressed in a C - like programming language and they can also detect hidden correlations between the features of the problem and use only those features that contain the most important information for the correct classification or patterns of the objective problem. In contrast to conventional AI methods, the proposed framework produces interpretable C-like rules while simultaneously performing implicit feature selection, enabling more explainable and efficient climate-prediction models.

The rest of this manuscript is organized as follows: the used dataset and the incorporated methods used in the conducted experiments are outlined in section 2, the experimental results are shown and discussed in section 3 and finally a detailed discussion is provided in section 4.

2. Materials and Methods

In this section, a detailed presentation of the datasets used as well as the machine learning techniques used in the experiments performed will be provided.

2.1. Impacts of Climate Change in Natural Disasters

In order to present the impacts of climate change in natural disasters, visual representations of the planet's average temperature are provided for the years 1881, 1995, and 2024. Visualization is from NASA Goddard's Scientific Visualization Studio: https://svs.gsfc.nasa.gov/5450

"This color-coded map shows changing global surface temperature anomalies. Normal temperatures are shown in white, higher than normal temperatures in red and lower than normal temperatures in blue".

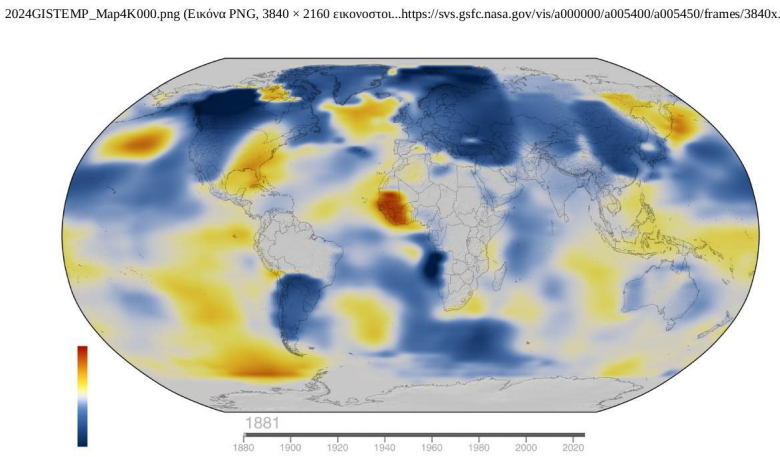


Figure 1. Global Temperature 1881

2024GISTEMP_Map4K685.png (Εικόνα PNG, 3840 × 2160 εικονοστοι...https://svs.gsfc.nasa.gov/vis/a000000/a005400/a005450/frames/3840x...

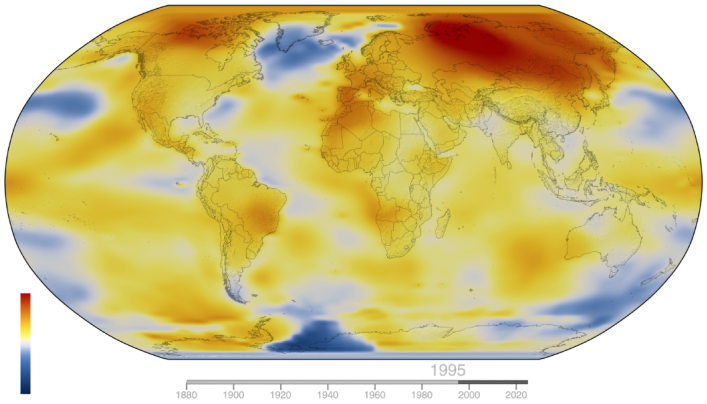


Figure 2. Global Temperature 1995

2024GISTEMP_Map4K899.png (Εικόνα PNG, 3840 × 2160 εικονοστοι...https://svs.gsfc.nasa.gov/vis/a000000/a005400/a005450/frames/3840x...

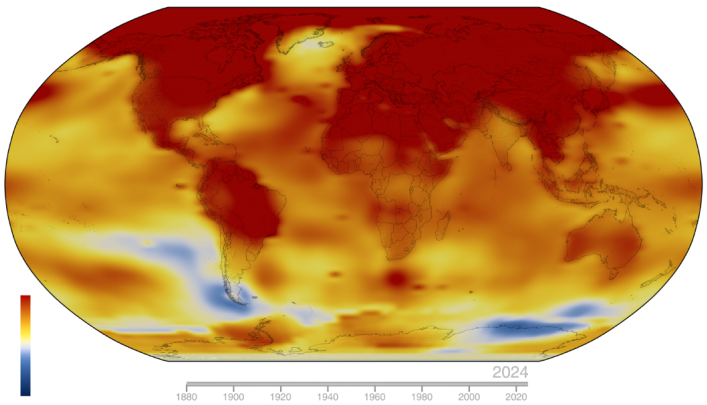


Figure 3. Global Temperature 2025

The 1.5°C target is highly ambitious and is likely to have been surpassed by 2025, nevertheless, it remains significant as a benchmark for minimizing, to the greatest extent possible, the impacts of natural disasters that are exacerbated by climate change. Subsequently, we will present numerical evidence illustrating how major natural disasters have become increasingly frequent and intense over time, as a result of the evolving impacts of climate change. Our data were retrieved from the (EM-DAT) Emergency Events Database [54] https://www.emdat.be/ on 9 November 2025.

From 1900 to 1994, a total of 5,423 major events were recorded, around the world categorized as follows :

Table 1. Hazards (1900 - 1994)

Hazards	Type	Events	Human Causalties	Type	Events	Human Causalties	Type	Events	Human Causalties	Total Causalties/ Hazard
Climatological	Drought	307	11.708,981	Wildfire	107	2.307				11.711,288
Hydrological	Flood	1.431	6.812,496	Landslide (wet)	280	43.587				6.856,083
Meteorological	Storm	1.789	1.151,310	Fog	1	4.000	Extreme Temperature	110	13.843	1.169,153
Geophysical	Earthquake	829	1.578,863	Volcanic Activity	126	85.144	Landslide (dry)	32	4.104	1.668,111
Biological	Epidemic	351	9.453,050	Infestation	61	(no data)				9.453,050

From the starting point of our research in 1995 up to 2025, the following 8.855 major events were recorded, around the world categorized as follows:

Table 2. Hazards (1995 - 2025)

Hazards	Type	Events	Human Causalties	Type	Events	Human Causalties	Type	Events	Human Causalties	Total Causalties/Hazard
	Drought	351	24.723	Wildfire	194	1.740	Glacial Lake Outburst Floods	8	478	26.941
Hydrological	Flood	3.576	160.394	Landslide (wet)	447	21.847				182.241
Meteorological	Storm	1.957	235.377	Fog	0	0	Extreme Temperature	533	352.179	587.556
Geophysical	Earthquake	647	605.378	Volcanic Activity	73	1.234	Landslide (dry)	10	280	606.892
Biological	Epidemic	1.025	163.994	Infestation	31	(no data)				163.994
Extra - Terrestrial	Impact	1	(no data)							

According to the available data, we conclude that natural disasters have increased over the past 30 years when compared with the preceding 94 years. More specifically, there has been a rise in droughts, floods, wildfires, landslides, storms, extreme temperatures, and epidemics. Geophysical events, by contrast, appear to have remained stable in number, given that the comparison is made against a period three times longer. In addition, two new categories have emerged.

2.2. The Dataset Employed

In this study, we utilized Historical Weather Data provided by Visual Crossing https://www.visualcrossing.com/, (last accessed on 08 December 2025), which aggregates information from multiple authoritative sources, including national meteorological services, global climate models, and satellite observations.In order to gain access to the historical data, we first completed the registration process. Subsequently, by selecting data access, we defined the region of interest and the desired date range. We then proceeded with the data retrieval process, with a limit of 1,000 records per day in order to remain within the free access tier. The overall procedure is straightforward, and the only limitation encountered was the daily cap of 1,000 data points.

2.3. The Dataset Description

We downloaded historical meteorological data covering the period from 1995 to 2025, spanning approximately 30 years, to support the needs of our climate change research. In addition, we selected regions across different geographical latitudes and longitudes in order to obtain a more comprehensive understanding of the areas most affected by the global increase in temperature. The selected regions of interest are: Congo, Saudi Arabia, Greece, Turkey, Reykjavik, Svalbard, and Cape Town. The climatic data for all regions span the period from January 1, 1995, to December 8, 2025. Specifically, Turkey provides the most complete dataset with 11,300 days of recordings, followed by Saudi Arabia with 11,299 days, Svalbard with 11,297 days, Greece with 11,291 days, Congo with 11,202 days, Iceland with 11,140 days, and finally Cape Town with 10,957 days. All regions have consistent 30 columns structure , 30+ years of data per region, and multiple climate zones, such as: Mediterranean (Greece & Turkey), Arctic (Iceland Reykjavik & Svalbard), Desert (Saudi Arabia), Tropical (Congo), Temperate (Cape Town). The differences in the number of data entries are due to gaps (missing values) present in the records obtained from Visual Crossing, which did not affect our results.

In summary, we report the maximum and minimum values of the average temperature for each region. For Greece, the highest recorded value was 36°C (96.8°F) on July 21, 2011, while the lowest was −26.33°C (-15.4°F) on January 27, 2003. In Turkey, the maximum temperature was recorded on July 30, 2000, at 41.05°C (105.9°F), while the minimum reached −22.94°C (−9.3°F) on February 16, 2006. In Saudi Arabia, the maximum temperature approached 48.2°C (118.9°F) on May 30, 2021, while the minimum reached −2.05°C (28.3°F) on January 17, 2008. In Cape Town, the maximum temperature reached 42°C (107.6°F) on February 3, 2015, while the minimum was 0.05°C (32.1°F), a value that occurred seven times:

Table 3. Table structure columns climate data

Number	Data
1	Temperature max (Fahrenheit)
2	Temperature min (Fahrenheit)
3	Feels like max (Fahrenheit)
4	Feels like min (Fahrenheit)
5	Feels like (Fahrenheit)
6	Dew
7	Humidity
8	Precipitation
9	Precipitation prob
10	Precipitation cover
11	Precipitation cover
12	Snow
13	Snow depth
14	Windgust
15	Wind speed
16	Wind dir
17	Sea level pressure
18	Cloud cover
19	Visibility
20	Solar radiation
21	Solar energy
22	Uvindex
23	Severe risk
24	Moon phase
25	Temp max (Celsius)
26	Temp min (Celsius)
27	Year
28	Month
29	Day
30	Class

Table 4. Mean Temperature Increased (1995 - 2025)

Number	Region	Mean Temperature increased
1	Turkey	1,92°C (3.46°F).
2	Reykjavik (Iceland)	1.41°C (2.54°F).
3	Greece	1.22°C (2.20°F).
4	Svalbard	0.93°C (1.67°F).
5	Cape Town	0.81°C (1.46°F).
6	Saudi Arabia	0.79°C (1.42°F).
7	Congo	0.10°C (0.18°F).

twice in July 1996, twice in 2000, and once in 2001, 2002, and 2003. In Reykjavik, Iceland, the maximum temperature was 25.7°C (78.3°F) on July 30, 2008, while the minimum reached −14.3°C (6.2°F) on July 3, 1998. In Congo, the maximum temperature reached 39.05°C (102.3°F) on February 22, 1995, while the minimum was 12.5°C (54.6°F) on June 7, 2012. In Svalbard, the maximum temperature reached 33.38°C (92.1°F) on July 20, 2022, while the minimum was −15.11°C (4.8°F) on January 5, 2003.

From these observations, it is striking that Greece recorded the lowest temperature −26.33°C (-15.4°F), followed by Turkey −22.94°C (−9.3°F) compared to regions located within the Arctic Circle, while the highest value was, as expected, observed in Saudi Arabia, a country characterized by a desert climate. In addition, over the period from January 1, 1995, to December 8, 2025, all seven regions under investigation exhibited an increase in mean temperature, albeit with varying magnitudes. Turkey recorded the highest rise at 1.92°C (3.46°F), followed by Reykjavik with 1.41°C (2.54°F) and Greece with 1.22°C (2.20°F). More moderate increases were observed in Svalbard (0.93°C; 1.67°F), Cape Town (0.81°C; 1.46°F), and Saudi Arabia (0.79°C; 1.42°F). Congo displayed the smallest change, with only 0.10°C (0.18°F). These findings highlight both regional disparities and the broader trend of warming consistent with the impacts of climate change.

After examining our dataset and identifying the variations in temperature across the seven regions under study, we proceeded with the classification of temperatures. To achieve this, the classification scheme required adaptation for each region individually. Specifically, we calculated the mean temperature for each region using the first five years of data, namely from 1995 to 2000. This mean value was employed as the baseline. Temperatures that were either lower or higher by up to 1°C (±1.8°F) relative to this baseline were assigned to Class 0, whereas temperatures exceeding the baseline by more than 1.1°C (≈2.0°F) were assigned to Class 1

The definition of the two classes was based on the environmental considerations discussed in our introduction, aligned with the Paris Agreement and the 1.5 °C climate target, as well as on our data-driven analyses and the class balance that we were able to achieve. Similarly, the temporal span of our dataset was selected to align with the climatological milestones commonly used to characterize the progression of climate change. For this reason, we adopted the period 1995–2025 as the analysis window across the different climatic regions. Furthermore, the definition of our classes was made with careful consideration, and the first class explicitly includes negative temperature anomalies, as it encompasses all temperature values from below zero up to +1 °C.

The rationale for selecting the initial five-year period was to establish a reference point unaffected by the extreme temperature fluctuations associated with climate change, given that our dataset spans from 1995 to 2025. This approach ensured that the classification was grounded in a stable climatic baseline, thereby enhancing the reliability of subsequent analyses.

Table 5. Baseline Class (1995 - 2000) / Region

Number	Region	Baseline for Class	Class 0 / total data	Class 1/ total data
1	Turkey	11.32	5.549	5.752
2	Reykjavik (Iceland)	4.715	5.681	5.620
3	Greece	9.068333333	5.629	5.662
4	Svalbard	8.966666667	5.957	5.345
5	Cape Town	16,.605	6.118	5.183
6	Saudi Arabia	26.88666667	5.531	5.767
7	Congo	25.6466667	7.536	3.666

2.4. The proposed method

The current work proposed the incorporation of a method that constructs classification rules, initially proposed in [55]. Furthermore, this method was implemented as a software recently [56]. The published software GENCLASS is an open-source software tool, publicly available on GitHub (https://github.com/itsoulos/GenClass/), where detailed installation instructions and usage examples are provided. The software is implemented in ANSI C++ and supports Python-formatted rule output. Its execution is controlled through a set of command-line parameters that define all key aspects of the method, including the number of chromosomes, number of generations, mutation rate, and other evolutionary settings. The implementation relies on OpenMP for multicore parallelization and uses Grammatical Evolution with a BNF grammar to generate human-readable classification rules.

The objective of the proposed method is to achieve high-accuracy classification of daily climate observations into temperature-change categories, while preserving interpretability through explicit rules. Grammatical Evolution produces concrete if-then programs over the available features, making nonlinear relationships transparent and directly inspectable. Empirically, GENCLASS demonstrates better performance over the considered baselines (MLP, RBF, SVM, and NNC) across all regions, achieving the lowest mean classification error (AVERAGE 1.66%) and consistently high precision/recall. Although strong deep sequence models for climate time series exist (e.g., LSTM/CDLSTM), the present study focuses on interpretable rule-induction and benchmarking against established conventional learners; a comprehensive comparison against deep sequence architectures under strictly time-ordered evaluation protocols is a natural extension for future work.

The proposed model, through the use of BNF grammar and Grammatical Evolution, is able to construct rules that exploit only the necessary features of the dataset under study. In this way, the model becomes more flexible and faster in decision-making, even when dealing with large datasets. Moreover, since it relies solely on the essential information for rule generation, it may exhibit better generalization capabilities compared to other machine learning models. The underlying grammar for this method is provided in Figure 4.

Figure 4. The underlying grammar for the method that produces classification programs using Grammatical Evolution.

```
<S>::=if(<BEXPR>) CLASS=0 else CLASS=1
<BEXPR>::=<XLIST><BOOLOP><EXPR>
         |!(<BEXPR>)
         |<XLIST><BOOLOP><EXPR>&<BEXPR>
         |<XLIST><BOOLOP><EXPR>|<BEXPR>
<BOOLOP>::= >
          | >=
          | <
          | <=
<EXPR>::= (<EXPR><BINARYOP><EXPR>)
        | <FUNCTION>(<EXPR>)
        | <TERMINAL>
<BINARYOP>::= +
            | -
            | *
            | /
<FUNCTION>::= sin | cos | exp | log
<TERMINAL>::= <XLIST>
            | <DIGITLIST>.<DIGITLIST>
            | (-<DIGITLIST>.<DIGITLIST>)
 <XLIST>::= x1 | x2 | ...|xD
<DIGITLIST>::=<DIGIT>
             | <DIGIT><DIGIT>
             | <DIGIT><DIGIT><DIGIT>
<DIGIT>::= 0 | 1 | 2 | 3 | 4 |5 |6 |7 |8 |9
```

The steps for this method have as follows:

1. **Initialization step**.
   (a) **Set** with $N_c$the number of chromosomes and with N_g the number of allowed generations.
   (b) **Set** the selection rate p_s and the mutation rate p_m.
   (c) **Initialize** each chromosome c_i, $i = 1, \ldots, N_c$ as a set of randomly selected integers.
   (d) **Set** $k = 0$, that defines the generation counter.
2. **Fitness calculation step**.
   (a) **For** $i = 1, \ldots, N_c$ **do**
      i. **Construct** with the procedure of Grammatical evolution and the incorporation of the BNF grammar of Figure 4 a classification program G_i that corresponds to the elements of chromosome c_i.
      ii. **Calculate** the associated fitness f_i value as

$$f_i = \sum_{j=1}^{M} \left(G_i(x_j) - t_j \right)^2 \tag{1}$$

The set $T = \{(x_1, t_1), (x_2, t_2), \ldots, (x_M, t_M)\}$ denotes the training set of the objective problem, where the vectors x_i stand for the input patterns and each value t_i represents the actual output for pattern x_i. The proposed method in practice uses the mean classification error on the training set as the fitness measure, since the output of the produced model is one of the available categories of the dataset under study.

(b) **End For**

3. **Application of Genetic Operations**.
   (a) **Selection procedure.** The remaining are substituted by chromosomes produced during crossover and mutation. The top $p_s \times N_c$ chromosomes are directly carried over to the next generation, while the rest are replaced by chromosomes generated through crossover and mutation.
   (b) **Crossover procedure.** This procedure produces new chromosomes from the current population. For every pair (c_1, c_2) of new chromosomes two chromosomes denoted as p_1 and p_2 are selected from the current population using tournament selection. The production of (c_1, c_2) will be performed using one - point crossover [57]. An example of this procedure is outlined in Figure 5.
   (c) **Mutation procedure**. Draw a random number $r \leq 1$ for each element of every chromosome. Alter randomly this element when $r \leq p_m$. The mutation procedure is depicted as a flowchart in Figure 6.
4. **Termination check step**.
   (a) **Set** $k = k + 1$
   (b) **If** $k < N_g$ then go to Fitness calculation step else go to Testing step.
5. **Testing step**.
   (a) **Get** the best chromosome c^* of the population with the lowest fitness value and create the corresponding classification program G^*.
   (b) **Apply** the classification program to the test set of the objective problem and calculate the corresponding test error.

The overall procedure is graphically outlined in Figure 7.

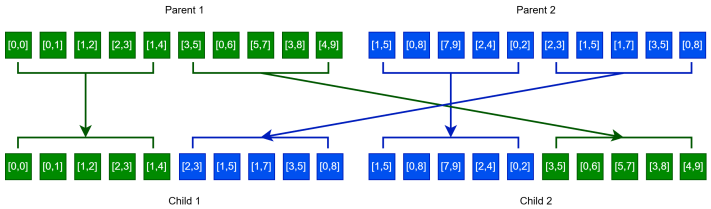


Figure 5. A graphical example of the one - point crossover.

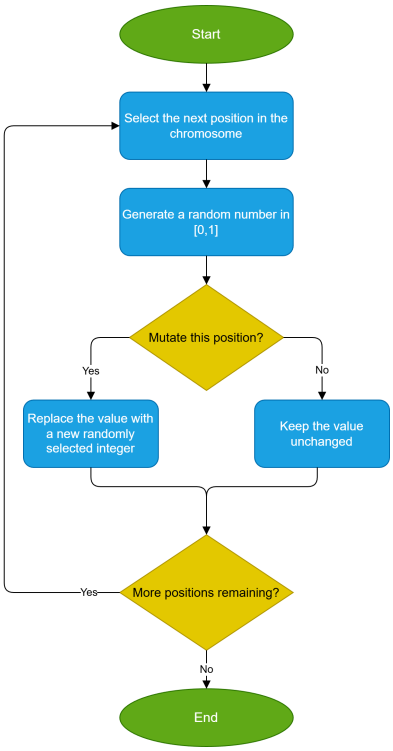


Figure 6. The flowchart of the mutation procedure.

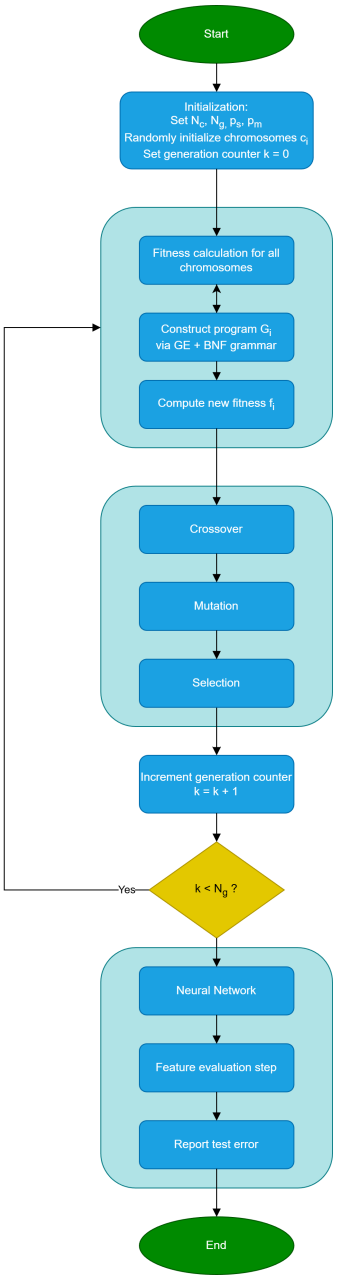


Figure 7. The flowchart of the proposed method.

3. Results

The software used in the experiments were coded in C++ programming language and the freely available software Optimus [58], available from https://github.com/itsoulos/GlobalOptimus/ (accessed on 25 December 2025), was incorporated for the optimization methods. The experimental results were validated using the ten - fold cross validation procedure and all the experiments were executed on a Debian Linux machine with 128GB of ram. The values for the experimental parameters are outlined in Table 6.

Table 6. Experimental settings.

PARAMETER	PURPOSE	VALUE
N_c	Number of chromosomes	500
N_g	Maximum number of generations	1000
p_s	Selection rate	0.10
p_m	Mutation rate	0.05
H	Number or weights for neural network	10

Table 7 summarizes the experimental results on datasets derived from various areas using a series of machine learning methods. The following notation is used in the experimental tables:

1. The column dataset denotes the objective dataset.
2. The column MLP represents the results derived from the application of an artificial neural network with 10 processing nodes. This network was trained using the BFGS variant of Powell [59] as the training method.
3. The column RBF denotes the application of a Radial Basis Function (RBF) network [60,61] to the related dataset. This network has 10 processing nodes.
4. The column SVM represents the application of the SVM method, using the implementation provided in the freely available library of LibSvm [62].
5. The column NNC stands for the usage of the Neural Network Construction method, initially described in [63]. The parameters used in the genetic algorithm for the NNC model are the same as in the case of the GENCLASS model.
6. The column GENCLASS indicates the incorporation of the proposed method for construction of classification rules.
7. The row average represents the average classification error for all used datasets.

Table 7 reports the percentage classification error obtained by five machine learning models (MLP, RBF, SVM, NNC, and the proposed GENCLASS) across seven region-specific climate-zone datasets (CAPETOWN, GREECE, ICELAND, KONGO, ARABIA, SVALBARD, TURKEY). Since the entries are error percentages, lower values indicate higher classification accuracy, while the final row (AVERAGE) summarizes the mean error over all datasets. The dominant outcome is the consistent superiority of GENCLASS, which achieves the lowest error on every dataset and the best overall average (1.66%). Importantly, this performance should be interpreted in light of the strong geographical and environmental heterogeneity represented by the selected regions, spanning a wide range of latitudes and longitudes and encompassing markedly different physical settings (coastal versus inland areas, mountainous terrain, forests, deserts, glaciers/sea ice, and distinct hemispheric seasonality). Such factors affect local temperature dynamics, variability, and seasonality and therefore shape the complexity of the classification patterns that each model must learn. For instance, KONGO lies close to the Equator (approximately within a few degrees north and south of 0° latitude, primarily at eastern longitudes in Central Africa) and is characterized by extensive tropical rainforest cover and high atmospheric moisture. In this dataset, Table 7 shows the highest errors for most methods and also the largest GENCLASS error (4.84%), suggesting a comparatively more challenging classification structure, plausibly due to weaker temperature seasonality and stronger interactions with humidity and cloud-related processes. In contrast, ARABIA (roughly 16°-32°N at eastern longitudes in Western Asia) is dominated by hot desert conditions, where land-surface properties, low cloudiness, and persistent aridity often yield clearer and more repeatable temperature-related signatures, accordingly, GENCLASS attains a very low error (0.73%) and NNC also performs strongly. In the Mediterranean/subtropical band, GREECE (approximately

35°-41°N, 20°-28°E) combines strong maritime influence, extensive coastline and island geography, and pronounced mountainous terrain, all of which introduce microclimatic variability, nevertheless, GENCLASS remains highly accurate (0.85%), indicating robust handling of nonlinear interactions linked to elevation, coastal effects, and seasonal transitions typical of Mediterranean climates. TURKEY (approximately 36°-42°N, 26°-45°E) is even more heterogeneous, with coastal zones influenced by surrounding seas and inland plateaus and mountain ranges exhibiting more continental conditions, despite this diversity, GENCLASS again yields the smallest error (0.63%), whereas conventional baselines such as MLP and SVM show substantially larger errors. At higher latitudes, ICELAND (approximately 63°-66°N, at western longitudes in the North Atlantic) is shaped by oceanic forcing, strong winds, glaciers, and complex volcanic topography, which can increase meteorological variability, still, GENCLASS keeps the error relatively low (2.47%) compared with the other methods. SVALBARD (approximately 74°-81°N at eastern longitudes in the Arctic Ocean) represents an Arctic archipelago with extensive glaciation/sea-ice influence and extreme photoperiodicity (polar night and midnight sun), conditions that can produce strong seasonal contrasts and circulation-driven variability, nevertheless, GENCLASS remains the most accurate method (1.11%), with NNC as the next best performer. Finally, CAPETOWN (approximately 34°S, 18°E) lies in the Southern Hemisphere and is typically described by a Mediterranean-type regime with strong coastal forcing and nearby mountainous relief that modulates winds and temperature variability, here too GENCLASS attains the lowest error (0.97%). The AVERAGE row confirms that, when aggregating across all these diverse geographic coordinates and environmental contexts, GENCLASS provides the most reliable generalization with the smallest mean error, while NNC consistently ranks second. Overall, the 7 supports not only that GENCLASS is more accurate, but also that it generalizes effectively across datasets derived from regions with fundamentally different latitude–longitude positioning and physical environments, suggesting improved capacity to capture nonlinear relationships and interactions induced by topography, land cover (forest/desert/ice), proximity to oceans, and regional seasonality.

Table 7. Experimental results on the provided datasets using a series of machine learning methods.

DATASET	MLP	RBF	SVM	NNC	GENCLASS
CAPETOWN	14.89%	12.43%	17.18%	2.00%	0.97%
GREECE	38.10%	4.34%	50.61%	6.15%	0.85%
ICELAND	36.87%	7.09%	49.79%	6.33%	2.47%
KONGO	26.52%	30.16%	32.71%	13.75%	4.84%
ARABIA	41.63%	7.62%	48.94%	4.91%	0.73%
SVALBARD	27.56%	23.30%	47.22%	5.07%	1.11%
TURKEY	33.53%	7.81%	49.10%	5.05%	0.63%
AVERAGE	**31.30%**	**13.25%**	**42.22%**	**6.18%**	**1.66%**

The results summarized in Figure 8 indicate that the comparison GENCLASS vs MLP yields $p = 0.0195$, implying a statistically significant difference ($p < 0.05$) and supporting the conclusion that GENCLASS outperforms MLP under the adopted experimental setting. The comparison GENCLASS vs RBF results in $p = 0.2492$, which is not significant (ns, $p > 0.05$), therefore, given the available datasets, there is insufficient statistical evidence to claim a systematic performance difference between GENCLASS and RBF at the 5% significance level. In contrast, GENCLASS vs SVM yields $p = 8.74 \times 10^{-5}$, an extremely small value ($p < 0.0001$), providing very strong statistical evidence that GENCLASS outperforms SVM on the evaluated datasets. The comparison GENCLASS vs NNC produces $p = 0.6583$ (ns), suggesting that no statistically significant difference can be established between GENCLASS and NNC. Finally, the overall Friedman test reports $p = 2.49 \times 10^{-5}$ ($p <$

0.0001), confirming that, when all models are jointly considered across all datasets, there are statistically significant performance differences and the methods cannot be regarded as equivalent. Overall, these findings substantiate that GENCLASS is decisively and statistically superior to SVM and significantly different from MLP, while differences relative to RBF and NNC are not statistically significant under the $p < 0.05$ criterion.

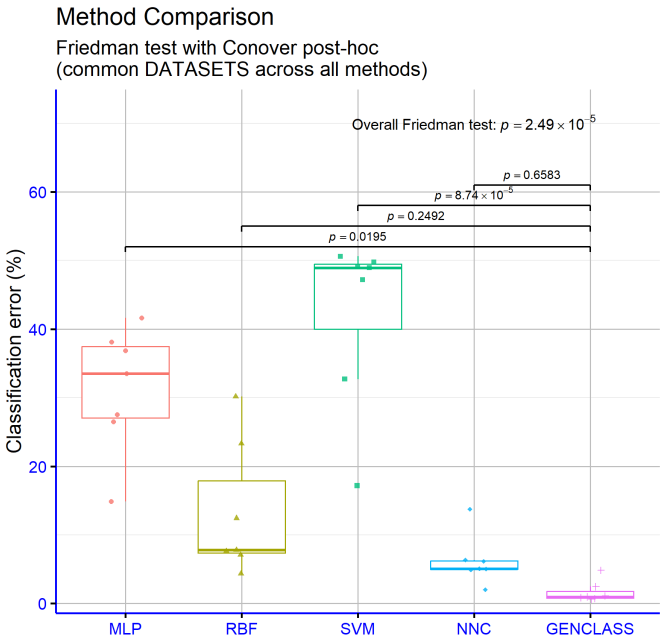


Figure 8. Statistical comparison for the results obtained by the application of various machine learning methods.

The dataset related to the KONGO country remains the most challenging dataset. A possible explanation for this outcome is that, as noted in the detailed Dataset Description in Section 2.2, the Congo region does not exhibit balanced class distributions compared to the other areas. This imbalance likely contributes to the slight deviation observed relative to the rest of our data.

The Figure 9 reports precision and recall for the three learning models (MLP, RBF, and GENCLASS) across the seven regions (ARABIA, CAPETOWN, GREECE, ICELAND, KONGO, SVALBARD, TURKEY). Precision reflects the purity of positive predictions (the fraction of predicted positives that are correct), whereas recall reflects the detection rate of the actual positives (the fraction of true positives that are correctly identified). GENCLASS consistently achieves the best performance on both metrics for all regions, exhibiting near-ceiling values with limited variation (precision approximately 0.950-0.996 and recall approximately 0.946-0.996). RBF is generally the second-best approach and reaches high values in several regions (e.g., GREECE/ICELAND/ARABIA/TURKEY around 0.92-0.96), but it degrades sharply in CAPETOWN and KONGO (precision 0.571 and 0.585, recall 0.735 and 0.644), which is visible as pronounced drops in the corresponding curve. MLP yields the weakest results overall, with precision approximately 0.576-0.711 and recall approximately 0.711-0.852, indicating substantially lower correctness of positive predictions and reduced ability to capture positive instances compared to the other methods. Overall, the figure provides clear empirical evidence that GENCLASS delivers the most robust and consistently strong classification behavior across regions while maintaining simultaneously high precision and high recall.

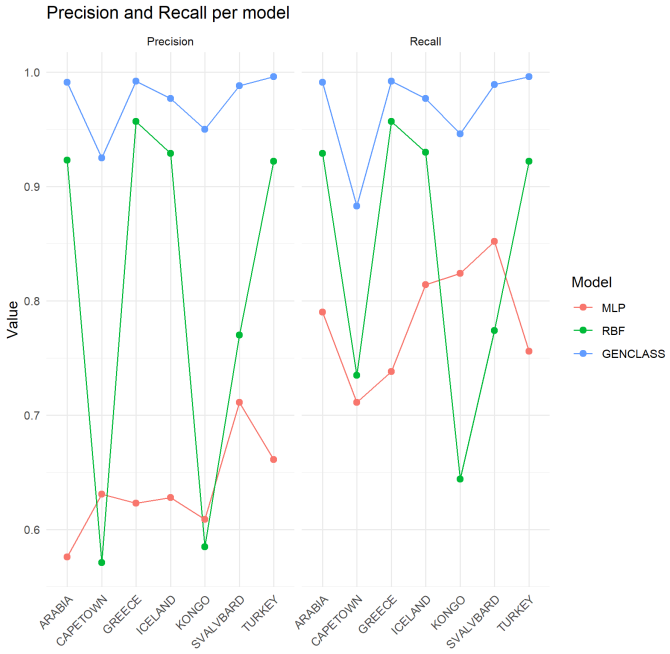


Figure 9. Precision and recall for three methods participated in the conducted experiments.

In addition, an indicative diagram showing the learning path for model GENCLASS and for KONGO dataset is presented in Figure 10.

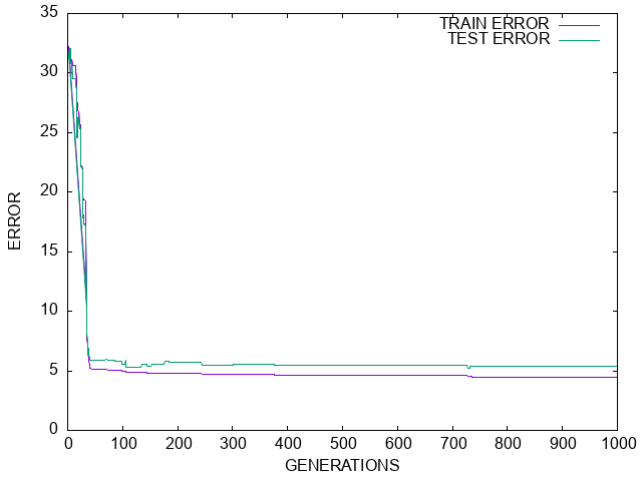


Figure 10. An indicative plot for the method GENCLASS and the dataset KONGO.

Moreover, Indicative plots for ROC and PR performance curves are outlined in Figures 11 and 12 respectively.

Figure 11. Indicative plots for the ROC curves, involving the RBF network, the NNC method and the GENCLASS method.

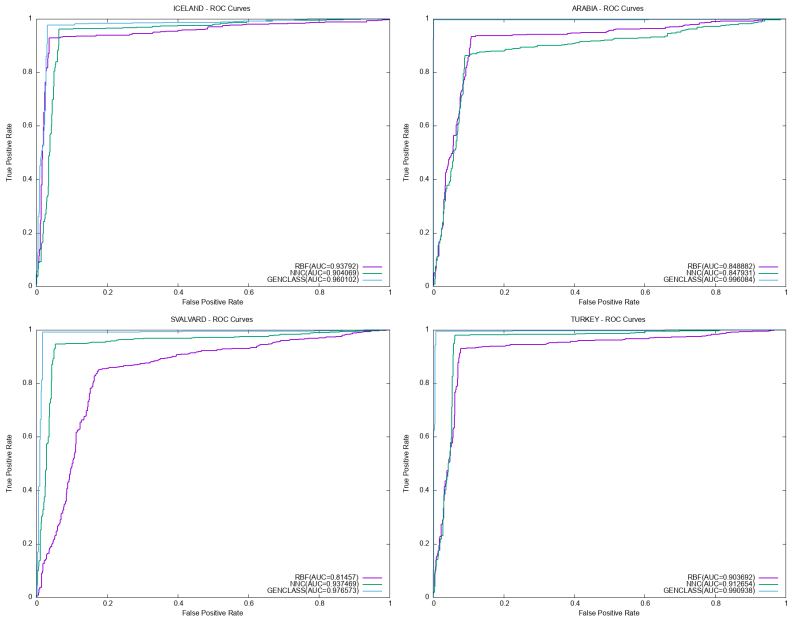
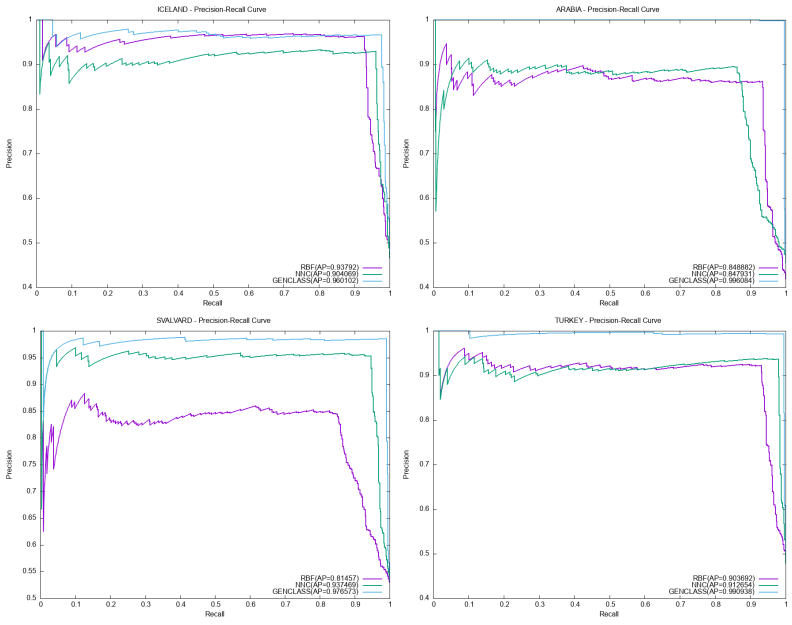


Figure 12. Indicative plots for the PR curves for the RBF network, the NNC method and the GENCLASS method.



As an example of rule production from the GENCLASS method, consider the program presented in Algorithm 1for the Greece dataset. As can be observed, the rule-generation method produces rules in a form that is understandable by humans (the above program can also be exported to actual C++ and Python code through the available software). Furthermore, for the effective creation of a classification program, it is not necessary to use all the features of the dataset under study.”

Algorithm 1 An example of classification program for the Greece dataset.

```
if (x26>=(((-61.4)+x3)/(-01.085))&x11<=x4&
    !(x1<(-58.007)|x21>((-561.6)*((-563.231)*((x28+x1)+(-77.456))))&
    x12>=(x5+(-47.804)))) CLASS=0.00
else
CLASS=1.00
```

3.1. Experiments with the number of chromosomes

Table 8 evaluates the sensitivity of the proposed model to the number of chromosomes N_c, using classification error percentages as the performance metric (lower is better). A clear and nearly monotonic improvement is observed as N_c increases from 50 to 1000: the mean error (AVERAGE row) drops from 3.21% to 1.46%, i.e., an absolute reduction of 1.75 percentage points corresponding to an approximate 54.5% relative decrease. Most of the gain is realized up to N_c=500 (3.21%→2.68%→2.17%→1.66%), whereas the additional improvement from 500 to 1000 is smaller (1.66%→1.46%, a 0.20 decrease), indicating diminishing returns for very large chromosome counts. At the dataset level, the largest relative reductions from N_c=50 to N_c=1000 occur for ARABIA (2.52%→0.54%, ~78.6%) and TURKEY (2.37%→0.57%, ~75.9%). In contrast, ICELAND exhibits a smaller but consistent overall reduction (3.22%→2.14%, ~33.5%) and a slight non-monotonic fluctuation at N_c=100 (3.27% vs. 3.22%), which is corrected for higher N_c. KONGO remains the most challenging dataset across all settings (8.86% at N_c=50 and 4.39% at N_c=1000), implying that increasing N_c substantially improves performance but does not fully remove the intrinsic difficulty associated with that region's data structure. Importantly, cross-dataset variability also decreases with N_c: the maximum error decreases from 8.86% to 4.39% and the range (max-min) shrinks from 7.08 to 3.85 percentage points, while the standard deviation of errors across datasets declines from about 2.54 to 1.40, suggesting more uniform generalization when using larger chromosome counts.

Table 8. Experimental results on the provided datasets using the proposed method and a series of values for the number of chromosomes N_c.

DATASET	$N_c = 50$	$N_c = 100$	$N_c = 200$	$N_c = 500$	$N_c = 1000$
CAPETOWN	1.85%	1.63%	1.40%	0.97%	0.78%
GREECE	1.78%	1.24%	1.13%	0.85%	0.83%
ICELAND	3.22%	3.27%	2.75%	2.47%	2.14%
KONGO	8.86%	8.18%	6.31%	4.84%	4.39%
ARABIA	2.52%	1.32%	1.26%	0.73%	0.54%
SVALBARD	1.89%	1.58%	1.30%	1.11%	0.94%
TURKEY	2.37%	1.57%	1.06%	0.63%	0.57%
AVERAGE	**3.21%**	**2.68%**	**2.17%**	**1.66%**	**1.46%**

In Figure 13 clearly illustrates that increasing the number of chromosomes leads to lower classification error across essentially all datasets, with the mean performance (AVERAGE, black dashed line) following a steady downward trend. The improvement is most pronounced from small N_c values up to approximately N_c=500, after which the error continues to decrease but at a slower rate, indicating diminishing returns for very large chromosome counts. Moreover, the dataset-specific curves become slightly more clustered as N_c increases, suggesting a more uniform behavior of the proposed model across regions.

Increasing the number of chromosomes will allow the genetic algorithm underlying the GENCLASS technique to perform a more thorough exploration of the problem's search space and to generate a larger number of classification rules. In this way, higher accuracy on the model's error in the test set can be achieved. Naturally, this will also lead to increased

training times; however, this drawback can be mitigated through the use of parallel Genetic Algorithms.

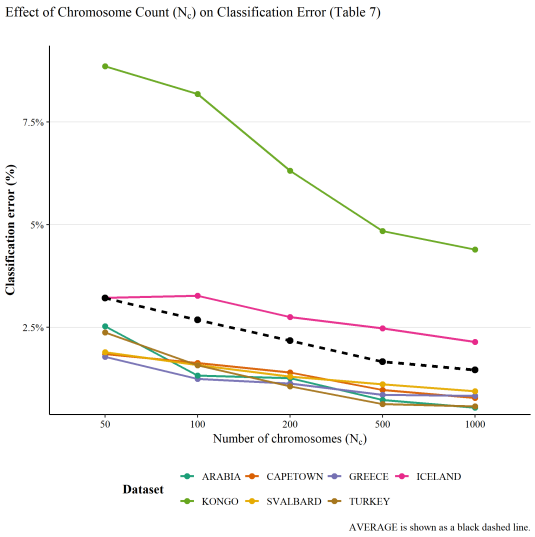


Figure 13. This figure shows the effect of changing the number of chromosomes on the learning error of the proposed technique for a series of datasets.

Also, the average execution time for the proposed method and a variety of values for the N_c parameter is outlined in Figure 14. As it was expected, the execution time increases rapidly as the number of chromosomes increases.

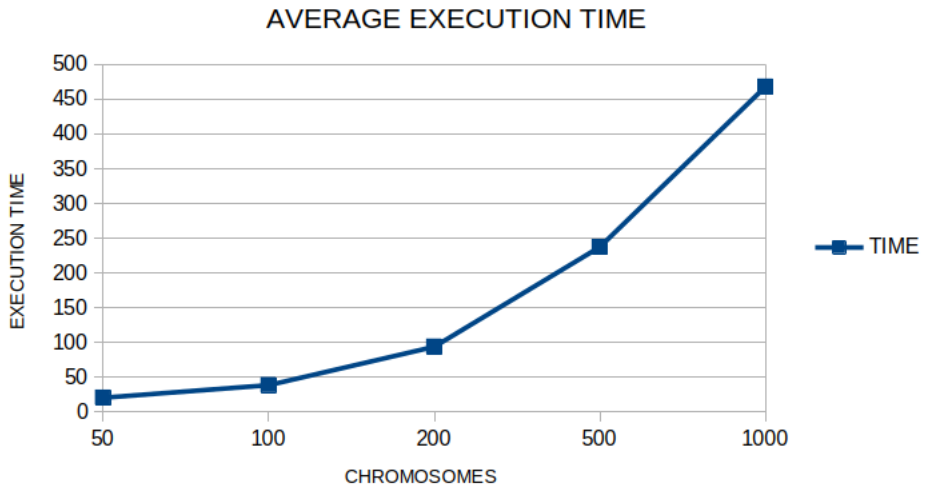


Figure 14. Average execution time for the proposed method and different values of parameter N_c.

3.2. Experiments with the number of generations

Table 9 investigates the effect of the number of generations N_g on the proposed model, using classification error percentages as the evaluation metric (lower is better). The overall trend is strongly improving with increasing N_g: the mean error (AVERAGE) decreases from 4.46% at N_g=50 to 1.45% at N_g=2000, corresponding to a 3.01 percentage-point reduction and an approximate 67.4% relative decrease. The gains are not evenly distributed across increments: 50→100 (-0.99), 100→200 (-0.82), 200→1000 (-0.98), and 1000→2000 (-0.21), indicating diminishing returns beyond 1000 generations, where additional computational effort yields smaller improvements in the mean error. At the dataset level, all regions improve substantially from N_g=50 to N_g=2000, with the largest relative reductions observed for

TURKEY (2.63%→0.37%, ~85.9%) and ARABIA (3.03%→0.63%, ~79.2%). KONGO also improves markedly (14.12%→4.18%, ~70.4%) but remains the most challenging dataset across all settings, consistently exhibiting the highest errors. ICELAND shows the smallest overall relative improvement (4.32%→2.58%, ~40.3%) and a slight degradation from N_g=1000 to N_g=2000 (2.47%→2.58%), suggesting that for certain datasets additional generations may introduce over-training effects and/or stochastic variability that partially offsets expected gains. The slight degradation in the ICELAND dataset between $N_g = 1000$ and $N_g = 2000$ can be attributed to stochastic fluctuations in the evolutionary search and dataset-specific characteristics that limit further improvement beyond a certain number of generations. Increasing N_g also reduces cross-dataset variability: the maximum error drops from 14.12% to 4.18% and the range (max–min) shrinks from 11.92 to 3.81 percentage points, while the standard deviation across datasets decreases from about 4.00 to 1.30, supporting a more uniform generalization profile at higher N_g.

Table 9. Experimental results on the provided datasets using the proposed method and a series of values for the number of generations N_g.

DATASET	$N_g = 50$	$N_g = 100$	$N_g = 200$	$N_g = 1000$	$N_g = 2000$
CAPETOWN	2.40%	1.96%	1.60%	0.97%	0.73%
GREECE	2.20%	1.74%	1.32%	0.85%	0.82%
ICELAND	4.32%	3.75%	3.05%	2.47%	2.58%
KONGO	14.12%	10.49%	7.91%	4.84%	4.18%
ARABIA	3.03%	2.35%	1.67%	0.73%	0.63%
SVALBARD	2.50%	1.98%	1.54%	1.11%	0.85%
TURKEY	2.63%	1.97%	1.38%	0.63%	0.37%
AVERAGE	**4.46%**	**3.46%**	**2.64%**	**1.66%**	**1.45%**

In Figure 15 shows that increasing the number of generations also yields a systematic reduction in classification error, with the mean curve (AVERAGE) improving markedly when moving from few to many generations. The largest gains are achieved up to roughly N_g=1000, while the transition from N_g=1000 to N_g=2000 exhibits a clear flattening of the curves, implying that additional computational effort produces smaller marginal improvements. Although minor dataset-specific fluctuations may occur at the final step, the overall pattern confirms that larger N_g enhances generalization and reduces error.

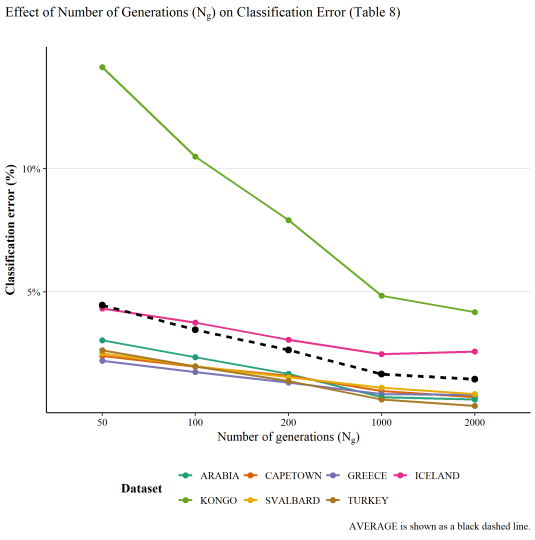


Figure 15. The impact of altering the number of generations to the classification error of the proposed method.

As shown here and according to established GA theory, increasing the number of generations allows the evolutionary process to explore a larger portion of the search space and refine candidate solutions over time, leading to more stable convergence and improved generalization across datasets.

3.3. Experiments with the RBF network

Another experiment was conducted where the RBF network was used as the classification model and the number of weights H was varied from 10 to 30 and the corresponding results are listed in Table 10.

Table 10. Experimental results using different number for the weights H of the RBF network.

DATASET	RBF $H = 10$	RBF $H = 15$	RBF $H = 20$	RBF $H = 25$	RBF $H = 30$	GENCLASS
CAPETOWN	12.43%	11.50%	10.77%	10.55%	10.41%	0.97%
GREECE	4.34%	3.34%	2.98%	9.06%	18.69%	0.85%
ICELAND	7.09%	4.79%	4.06%	3.33%	3.13%	2.47%
KONGO	30.16%	17.59%	19.93%	21.69%	21.21%	4.84%
ARABIA	7.62%	4.35%	3.40%	2.96%	2.81%	0.73%
SVALVARD	23.30%	9.65%	7.74%	6.92%	6.44%	1.11%
TURKEY	7.81%	7.00%	7.54%	6.77%	7.87%	0.63%
AVERAGE	**13.25%**	**8.32%**	**8.06%**	**8.75%**	**10.08%**	**1.66%**

The Figure 16 is a heatmap of classification error (Error %) for the RBF model under different numbers of hidden nodes H, together with a GENCLASS reference column. Each row corresponds to one region/dataset (CAPETOWN, GREECE, ICELAND, KONGO, ARABIA, SVALBARD, TURKEY), while the last row (AVERAGE) reports the mean error across all regions. Each column represents a different RBF configuration ($H = 10$, 15, 20, 25, 30), and the final column reports GENCLASS. The numerical error percentages are printed inside each cell, and the color intensity encodes the magnitude of the error (darker color indicates lower error). The heatmap shows that increasing H often improves RBF performance initially, but the trend is not monotonic and degradation is observed for larger H in some regions, which is consistent with overfitting; for example, in GREECE the error remains low up to $H = 20$ but increases sharply at $H = 25$ and $H = 30$, while in KONGO the best performance occurs at a smaller H and then worsens as H grows. In the AVERAGE row, the best mean RBF performance occurs around $H = 20$, whereas larger settings ($H = 25, H = 30$) lead to higher mean error, highlighting a saturation regime and diminishing returns. Across all regions, GENCLASS achieves markedly lower error than any tested RBF setting, indicating a consistently stronger reference performance against which RBF sensitivity to hidden-layer size can be assessed. All values in the figure are taken directly from the raw Excel table (without formatting) and visually summarize both the initial gains from increasing H and the potential performance drop when H becomes excessively large.

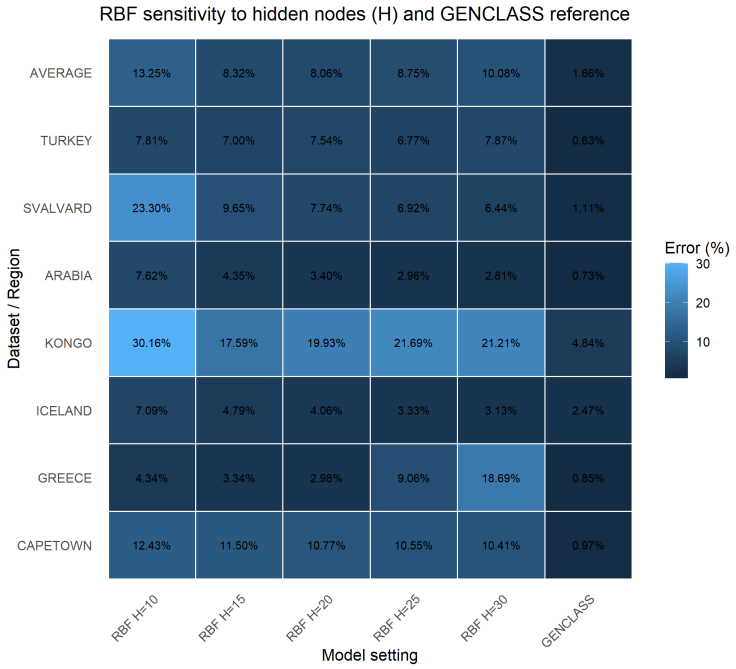


Figure 16. RBF Sensitivity to the number of hidden nodes.

4. Conclusions

The present study provides empirical evidence that daily temperature time series (1995-2025) across seven regions spanning distinct climate zones exhibit a consistent warming signal, albeit with pronounced spatial heterogeneity. Mean temperature increases are not uniform: Turkey shows the strongest rise, followed by Reykjavik and Greece, whereas the Congo region displays only marginal changes. This spatial contrast underscores that large-scale warming manifests differently depending on geography, seasonality, maritime influence, topography, and local atmospheric processes (e.g., humidity and cloud-related dynamics). In methodological terms, the adoption of a region-specific baseline derived from the early 1995-2000 period proved effective for defining comparable classes across heterogeneous climates, supporting relatively balanced class distributions and reducing the risk that the labeling scheme inherits the already intensified warming of recent years.

From a modeling perspective, Grammatical Evolution with rule-based construction (GENCLASS) demonstrates robust and systematic superiority over the considered baselines (MLP, RBF, SVM, and NNC) across all datasets, achieving the lowest average classification error and indicating strong cross-regional generalization. Importantly, the advantage is not limited to the mean performance but extends to consistency across markedly different environmental regimes, which is critical for transferability in climate-related applications. Statistical validation further corroborates that the observed differences are unlikely to be due to chance: the overall Friedman test is highly significant ($p = 2.49 \times 10^{-5}$). Pairwise comparisons indicate that GENCLASS significantly outperforms SVM and differs significantly from MLP, while differences relative to RBF and NNC are not statistically significant at the 0.05 level. This pattern is consistent with GENCLASS remaining competitive against strong conventional learners while avoiding the pronounced degradation that some methods exhibit on nonlinear and non-stationary climate time series.

Sensitivity analyses with respect to key evolutionary parameters strengthen the methodological credibility of the proposed approach. Increasing the number of chromosomes (N_c) and generations (N_g) yields systematic error reductions, with clear diminishing

returns beyond approximately $N_c \approx 500$ and $N_g \approx 1000$, suggesting that improvements stem from enhanced search efficacy rather than incidental fluctuations. Moreover, the persistent relative difficulty of the KONGO dataset across all settings is informative: in regions with weaker temperature seasonality and stronger coupling between temperature and moisture/cloud processes, a binary temperature-deviation labeling scheme may be intrinsically harder to learn, motivating either feature enrichment or a revised labeling strategy that better reflects the underlying physical dynamics.

The parameter sensitivity analysis indicates that increasing the number of chromosomes (N_c) and the number of generations (N_g) leads to systematic reductions in error, while the marginal gains progressively shrink beyond a moderate range of values, revealing clear diminishing returns. This pattern is expected in genetic and evolutionary computation: enlarging the computational budget (via larger populations and/or longer runs) typically improves exploration and convergence at first, but eventually enters a saturation regime once the search operates in a sufficiently good region of the solution space. Similar observations and interpretations have been reported in the literature on parameter setting/control in evolutionary algorithms, as well as in broader empirical studies of hyper-parameter spaces that identify wide viable regions and limited additional benefits beyond certain cost levels [64–67]. This contextualization supports reproducibility by documenting that the selected configuration lies within a stable-performance regime without requiring excessive computational overhead [68].

The effect of the number of generations N_g is reported in Section 3.2 (Table 9 and Figure 15). Error decreases strongly as N_g increases up to 1000, while the step from $N_g = 1000$ to $N_g = 2000$ exhibits a clear flattening (diminishing returns): the AVERAGE error drops from 1.66% $(N_g = 1000)$ to 1.45% $(N_g = 2000)$, i.e., only by 0.21 percentage points. For reproducibility and a balanced trade-off between performance and computational effort, $N_g = 1000$ is recommended as a practical default, which is also the value adopted in the main experimental settings (Table 6). This recommendation is consistent with the algorithm definition in Section 2.4, where N_g acts as the maximum number of generations in the termination condition, therefore, computational effort increases approximately proportionally with the number of generations (for fixed N_c). Optionally, an early-stopping rule (e.g., stopping when the best fitness does not improve for W consecutive generations) can be used to further balance cost and benefit, while retaining $N_g = 1000$ as the default configuration.

Despite the strong results, several limitations delineate the scope of generalization. First, the data originate from an aggregated provider and may embed temporal inconsistencies, missingness patterns, or changes in upstream data synthesis over time. Second, evaluation is conducted using classification error under a binary formulation (Class 0/1), which necessarily simplifies a multi-factor climate system and does not quantify predictive uncertainty. Third, careful time-series-aware evaluation is essential to preclude information leakage from later periods into earlier ones. Fourth, applying a uniform absolute threshold (±1°C and >1.1°C) for class definition while practical may not be climatically equivalent across tropical versus polar regimes, where variability and seasonal structure differ qualitatively.

Future work can extend this research along four complementary directions. First, geographic and climatic coverage should be broadened by incorporating additional regions per climate zone, explicitly addressing microclimates (island and mountainous settings), and systematically revisiting Antarctica through alternative data sources and missing-data strategies. Second, the representation of the climate system should be enriched by integrating additional predictors (humidity, pressure, wind, radiation, drought indices, and teleconnection-related indicators), enabling the learned rules to capture drivers rather

than only temperature outcomes. Third, the task formulation should be expanded beyond binary classification toward multi-class regimes and/or regression-based forecasting with explicit horizons (10/20/30 years) and uncertainty quantification, thereby improving the operational relevance of the results. Fourth, experimental protocols should be strengthened with time-aware cross-validation, decade-wise stability checks, and broader benchmarking against modern hybrid/deep time-series models, while retaining the interpretability advantages of rule-based structures. Ultimately, the most impactful direction is a unified framework that couples predictive accuracy with statistical rigor and interpretable rules, ensuring that the resulting climate inferences are not only accurate but also scientifically explainable and actionable for adaptation planning.

Author Contributions: C.K., V.C. and I.G.T. conceived of the idea and the methodology, and C.K. and V.C. implemented the corresponding software. C.K. conducted the experiments, employing objective functions as test cases, and provided the comparative experiments. V.C. performed the necessary statistical tests. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been financed by the European Union : Next Generation EU through the Program Greece 2.0 National Recovery and Resilience Plan , under the call RESEARCH-CREATE-INNOVATE, project name “iCREW: Intelligent small craft simulator for advanced crew training using Virtual Reality techniques" (project code:TAEDK-06195).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aeschylus, 472 B.C.The Persians. Place and date of edition used as base for this ebook: London: George Allen & Unwin, 1939. Available from:https://gutenberg.ca/ebooks/murrayaeschylus-persians/murrayaeschylus-persians-00-h.html
2. Aeschylus, 472 B.C. Persians. Richer Resources Publications, Arlington, Virginia, USA, ISBN 978-1-935238-57-7. Available from: https://johnstoniatexts.x10host.com/aeschylus/persianshtml.html
3. United Nations. Climate Action. Available from: https://www.un.org/en/climatechange/what-is-climate-change (accessed on 7 Novemer 2025)
4. United Nations. (1992) Climate Change. United Nations Framework Convention on Climate Change (UNFCCC). Rio de Janeiro, Brazil 3 - 14 June 1992. Available from: https://unfccc.int/process-and-meetings/united-nations-framework-convention-on-climate-change (accessed on 7 Novemer 2025)
5. United Nations. (1997) Climate Change.The Kyoto Protocol, 11 December 1997. Available from: https://unfccc.int/process-and-meetings/the-kyoto-protocol (accessed on 7 Novemer 2025)
6. Hallegatte, S. (2014). Natural disasters and climate change. Cham: Springer International Publishing. ISBN 978-3-319-08932-4 ISBN 978-3-319-08933-1 (eBook).
7. Kopitsa, C., Tsoulos, I. G., Miltiadous, A., & Charilogis, V. (2025). Predicting the Forest Fire Duration Enriched with Meteorological Data Using Feature Construction Techniques. Symmetry, 17(11), 1785.
8. United Nations Environment Programme. (2022) Number of Wildfires to Rise by 50 per Cent by 2100 and Governments Are Not Prepared, Experts Warn. 23 February 2022. Available online: https://www.unep.org/news-and-stories/press-release/number-wildfires-rise-50-cent-2100-and-governments-are-not-prepared (accessed on 7 Novemer 2025)
9. United Nations. (2015) Climate Action. The Paris Agreement. Available from: https://www.un.org/en/climatechange/paris-agreement (accessed on 8 November 2025).
10. NASA. 2024. Global Temperature - Earth Indicator. Available from: https://science.nasa.gov/earth/explore/earth-indicators/global-temperature/ (accessed on 8 November 2025).
11. Arrhenius, S. (1896). On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground. Philosophical Magazine and Journal of Science, Series 5, Volume 41, April 1896, pages 237-276. A
12. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Bengio, Y. (2022). Tackling climate change with machine learning. ACM Computing Surveys (CSUR), 55(2), 1-96.

13. Koulelis, P. P., Proutsos, N., Solomou, A. D., Avramidou, E. V., Malliarou, E., Athanasiou, M., ... & Petrakis, P. V. (2023). Effects of climate change on Greek forests: A review. Atmosphere, 14(7), 1155.
14. Rovithakis, A., Grillakis, M. G., Seiradakis, K. D., Giannakopoulos, C., Karali, A., Field, R., ... & Voulgarakis, A. (2022). Future climate change impact on wildfire danger over the Mediterranean: the case of Greece. Environmental Research Letters, 17(4), 045022.
15. Giannakopoulos, C., Kostopoulou, E., Varotsos, K. V., Tziotziou, K., & Plitharas, A. (2011). An integrated assessment of climate change impacts for Greece in the near future. Regional Environmental Change, 11(4), 829-843.
16. Knox, J. C. (2000). Sensitivity of modern and Holocene floods to climate change. Quaternary Science Reviews, 19(1-5), 439-457.
17. Milly, P. C. D., Wetherald, R. T., Dunne, K. A., & Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. Nature, 415(6871), 514-517.
18. Bronstert, A. (2003). Floods and climate change: interactions and impacts. Risk Analysis: An International Journal, 23(3), 545-557.
19. Cook, B. I., Mankin, J. S., & Anchukaitis, K. J. (2018). Climate change and drought: From past to future. Current Climate Change Reports, 4(2), 164-179.
20. Mukherjee, S., Mishra, A., & Trenberth, K. E. (2018). Climate change and drought: a perspective on drought indices. Current climate change reports, 4(2), 145-163.
21. Yuan, X., Wang, Y., Ji, P., Wu, P., Sheffield, J., & Otkin, J. A. (2023). A global transition to flash droughts under climate change. Science, 380(6641), 187-191.
22. Kretschmer, M., Runge, J., & Coumou, D. (2017). Early prediction of extreme stratospheric polar vortex states based on causal precursors. Geophysical research letters, 44(16), 8592-8600.
23. Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., & Voulgarakis, A. (2018). Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. Environmental Research Letters, 13(10), 104016.
24. Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters, 14(12), 124007.
25. Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., & Voulgarakis, A. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. npj Climate and Atmospheric Science, 3(1), 44.
26. Haggag, M., Siam, A. S., El-Dakhakhni, W., Coulibaly, P., & Hassini, E. (2021). A deep learning model for predicting climate-induced disasters. Natural Hazards, 107(1), 1009-1034.
27. Haq, M. A. (2022). CDLSTM: A novel model for climate change forecasting. Computers, Materials & Continua, 71(2).
28. Slater, L. J., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., ... & Zappa, M. (2023). Hybrid forecasting: blending climate predictions with AI models. Hydrology and earth system sciences, 27(9), 1865-1889.
29. Bist, A. S., Rawat, B., Joshi, Y., Aini, Q., Santoso, N. P. L., & Kusumawardhani, D. A. R. (2024, August). Harnessing Deep Learning for Accurate Climate Change Predictions. In 2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT) (pp. 1-6). IEEE.
30. GAUTAM, A., AJMERA, R., DHARAMDASANI, D. K., SRIVASTAVA, S., & JOHARI, A. (2025). IMPROVING CLIMATE CHANGE PREDICTIONS USING TIME SERIES ANALYSIS AND DEEP LEARNING. Global & Stochastic Analysis, 12(4).
31. O'Neill, M., & Ryan, C. (2002). Grammatical evolution. IEEE Transactions on Evolutionary Computation, 5(4), 349-358.
32. Kramer, O. (2017). Genetic algorithms. In Genetic algorithm essentials (pp. 11-19). Cham: Springer International Publishing.
33. Backus, J. W. (1959). The syntax and the semantics of the proposed international algebraic language of the Zurich ACM-GAMM Conference. In ICIP Proceedings (pp. 125-132).
34. Ryan, C., Collins, J. J., & Neill, M. O. (1998, April). Grammatical evolution: Evolving programs for an arbitrary language. In European conference on genetic programming (pp. 83-96). Berlin, Heidelberg: Springer Berlin Heidelberg.
35. O'Neill, M., & Ryan, C. (1999, May). Evolving multi-line compilable C programs. In European Conference on Genetic Programming (pp. 83-92). Berlin, Heidelberg: Springer Berlin Heidelberg.
36. Brabazon, A., & O'Neill, M. (2006). Credit classification using grammatical evolution. Informatica, 30(3).
37. Şen, S., & Clark, J. A. (2009, March). A grammatical evolution approach to intrusion detection on mobile ad hoc networks. In Proceedings of the second ACM conference on Wireless network security (pp. 95-102).
38. Chen, L., Tan, C. H., Kao, S. J., & Wang, T. S. (2008). Improvement of remote monitoring on water quality in a subtropical reservoir by incorporating grammatical evolution with parallel genetic algorithms into satellite imagery. Water Research, 42(1-2), 296-306.
39. Hidalgo, J. I., Colmenar, J. M., Risco-Martin, J. L., Cuesta-Infante, A., Maqueda, E., Botella, M., & Rubio, J. A. (2014). Modeling glycemia in humans by means of grammatical evolution. Applied Soft Computing, 20, 40-53.
40. Tavares, J., & Pereira, F. B. (2012, April). Automatic design of ant algorithms with grammatical evolution. In European conference on genetic programming (pp. 206-217). Berlin, Heidelberg: Springer Berlin Heidelberg.
41. Zapater, M., Risco-Martín, J. L., Arroba, P., Ayala, J. L., Moya, J. M., & Hermida, R. (2016). Runtime data center temperature prediction using grammatical evolution techniques. Applied Soft Computing, 49, 94-107.

42. Ryan, C., O'Neill, M., & Collins, J. J. (1998, June). Grammatical evolution: Solving trigonometric identities. In proceedings of Mendel (Vol. 98, p. 4th). Brno, Czech Republic: Technical University of Brno, Faculty of Mechanical Engineering.
43. de la Puente, A. O., Alfonso, R. S., & Moreno, M. A. (2002, June). Automatic composition of music by means of grammatical evolution. In Proceedings of the 2002 conference on APL: array processing languages: lore, problems, and applications (pp. 148-155).
44. De Campos, L. M. L., de Oliveira, R. C. L., & Roisenberg, M. (2016). Optimization of neural networks through grammatical evolution and a genetic algorithm. Expert Systems with Applications, 56, 368-384.
45. Soltanian, K., Ebnenasir, A., & Afsharchi, M. (2022). Modular grammatical evolution for the generation of artificial neural networks. Evolutionary computation, 30(2), 291-327.
46. Dempsey, I., O'Neill, M., & Brabazon, A. (2007). Constant creation in grammatical evolution. International Journal of Innovative Computing and Applications, 1(1), 23-38.
47. Galván-López, E., Swafford, J. M., O'Neill, M., & Brabazon, A. (2010, April). Evolving a ms. pacman controller using grammatical evolution. In European Conference on the Applications of Evolutionary Computation (pp. 161-170). Berlin, Heidelberg: Springer Berlin Heidelberg.
48. Shaker, N., Nicolau, M., Yannakakis, G. N., Togelius, J., & O'neill, M. (2012, September). Evolving levels for super mario bros using grammatical evolution. In 2012 IEEE Conference on Computational Intelligence and Games (CIG) (pp. 304-311). IEEE.
49. Martínez-Rodríguez, D., Colmenar, J. M., Hidalgo, J. I., Villanueva Micó, R. J., & Salcedo-Sanz, S. (2020). Particle swarm grammatical evolution for energy demand estimation. Energy Science & Engineering, 8(4), 1068-1079.
50. Sabar, N. R., Ayob, M., Kendall, G., & Qu, R. (2013). Grammatical evolution hyper-heuristic for combinatorial optimization problems. IEEE Transactions on Evolutionary Computation, 17(6), 840-861.
51. Ryan, C., Kshirsagar, M., Vaidya, G., Cunningham, A., & Sivaraman, R. (2022). Design of a cryptographically secure pseudo random number generator with grammatical evolution. Scientific reports, 12(1), 8602.
52. Pereira, P. J., Cortez, P., & Mendes, R. (2021). Multi-objective grammatical evolution of decision trees for mobile marketing user conversion prediction. Expert Systems with Applications, 168, 114287.
53. Castejón, F., & Carmona, E. J. (2018). Automatic design of analog electronic circuits using grammatical evolution. Applied Soft Computing, 62, 1003-1018.
54. EM-DATA. The International Disaster Database. Emergency Events Database. Available from: https://www.emdat.be/ (accessed on 9 November 2025).
55. Tsoulos, I.G. Creating classification rules using grammatical evolution. Int. J. Comput. Intell. Stud. 2020, 9, 161–171.
56. Anastasopoulos, N.; Tsoulos, I.G.; Tzallas, A. GenClass: A parallel tool for data classification based on Grammatical Evolution. SoftwareX 2021, 16, 100830.
57. Poli, R., & Langdon, W. B. (1998). Genetic programming with one-point crossover (pp. 180-189). Springer London.
58. Tsoulos, I. G., Charilogis, V., Kyrou, G., Stavrou, V. N., & Tzallas, A. (2025). OPTIMUS: A Multidimensional Global Optimization Package. Journal of Open Source Software, 10(108), 7584.
59. M.J.D Powell, A Tolerant Algorithm for Linearly Constrained Optimization Calculations, Mathematical Programming **45**, pp. 547-566, 1989.
60. J. Park and I. W. Sandberg, Universal Approximation Using Radial-Basis-Function Networks, Neural Computation **3**, pp. 246-257, 1991.
61. G.A. Montazer, D. Giveki, M. Karami, H. Rastegar, Radial basis function neural networks: A review. Comput. Rev. J **1**, pp. 52-74, 2018.
62. Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 1-27.
63. Tsoulos, I., Gavrilis, D., & Glavas, E. (2008). Neural network construction and training using grammatical evolution. Neurocomputing, 72(1-3), 269-277.
64. Eiben, A. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. IEEE Transactions on Evolutionary Computation, 3(2), 124–141.
65. De Jong, K. A. (2007). Parameter setting in EAs: A 30 year perspective. In F. G. Lobo, C. F. Lima, & Z. Michalewicz (Eds.), Parameter Setting in Evolutionary Algorithms (pp. 1–18). Springer.
66. Sipper, M., Fu, W., Ahuja, K., & Moore, J. H. (2018). Investigating the parameter space of evolutionary algorithms. BioData Mining, 11, 2.
67. O'Neill, M., Brabazon, A., Nicolau, M., McGarraghy, S., & Keenan, P. (Eds.). (2018). The Handbook of Grammatical Evolution. Springer.
68. Mills, K. L., Filliben, J. J., & Haines, A. (2014). Determining relative importance and best settings for genetic algorithm control parameters. Evolutionary Computation, 23(2), 309–342.