JOURNAL OF DATA SCIENCE | Vol.2024:30
eISSN:2805-5160

Predictive Analytics in Genetic Engineering as an Optimization Problem

Okewu Emmanuel[1] and Okewu Kehinde Bukola[2]

[1]Centre for Information Technology and Systems, University of Lagos, Nigeria
[2]Department of Plant Science, Federal University of Technology, Lagos, Nigeria

Email: eokewu@unilag.edu.ng

Abstract

In genetic engineering, developing a breed with a desired trait is a search and optimization problem that sometimes requires many generations of field and laboratory experiments for an optimal solution to be found. The nature of the problem requires that a stochastic optimization algorithm be applied in the metaheuristic search rather than using a deterministic or mathematical approach. In the search for drought-tolerant cowpea, this study applied a genetic algorithm as a predictive analytics tool in the genetic engineering of three native cowpea landraces *(Dan muzakkari, Gidigiwa, and Dan mesera)* selected from Northern Nigeria (specifically from Kontagora in Niger State of Nigeria). The three cowpea species were subjected to mutagenic treatments using gamma irradiation and Ethyl Methane Sulphonate (EMS). Doses applied include 200, 400, 600, and 800 Gray of gamma irradiation and 0.372% v/v of EMS. Both treated and untreated cowpea landraces were planted and observed. Mutation-induced breeding aims to deepen the drought-tolerant trait of the cowpea mutants to survive conditions in drought-prone Northern Nigeria. The statistical analysis of the agro-morphological and yield parameters of the first mutant generation (M1 generation) indicates that mutagenic treatments have a positive impact on both the yield and the survival of the three landraces as all the treated landraces yielded better than the control, particularly the treatments combination of 600gray and 372% v/v of EMS. Also, the predictive outcomes of the computational simulation that was implemented in Python programming indicate that these local cultivars are developing drought-tolerant genetic variability. For the three computational experiments, the stochastic optimizer (genetic algorithm) converged at the 9412th, 9717th, and 14338th generations respectively. Such predictive analytics information is useful for guiding decision-making by researchers and breeders in the crop improvement program.

Keywords

Constrained stochastic optimization, Cowpea, Genetic mutation, Stochastic optimizer, Sustainability

Introduction

The search for an optimal candidate species with the desired trait in crop improvement programs, using either natural breeding or genetic engineering, is a metaheuristic and optimization search that involves many generations to achieve. Because the search problem is not deterministic, applying the traditional mathematical approach to predict when an optimal solution would be obtained is difficult. Hence, computational algorithms are applied.

Submission: 27 June 2024; **Acceptance:** 4 August 2024

Developing crops with required traits is practically more achievable using genetic engineering than natural breeding (W. Yali, 2022). This is because it is easier to modify the genetic make-up of crops in a controlled setting that uses genetic engineering techniques like mutation-induction than relying on natural breeding whose outcomes are not certain let alone predicting the time the desired genotypic trait would be achieved (M. Antoniou, 2021) (J. Muthuvel., et.al., 2021). In any case, the genetic mutation process also suffers some degree of uncertainty as a result of biotic and abiotic factors (E. Okewu, et.al., 2023). Biotic factors such as pests, insects, etc., and abiotic factors like rain, light, and wind can impact the outcomes of mutation-induced breeding in terms of parameters like time and yield.

A system such as the crop improvement system whose outcomes are determined by uncertainties is a stochastic system (S.M. Kossivi., 2023) (M. Mohamed., 2023). The focus of breeders and researchers in such a system is to get the best possible solution within the context of the uncertainties, hence the problem being resolved is a stochastic optimization problem. Though there have been several research on using genetic engineering to induce desired traits in crops globally, reports on efforts to induce cowpeas with desired drought-tolerant traits so as to survive conditions in drought-prone Northern Nigeria have been scanty. This article reports on the genetic mutation of three cultivars of cowpea from Northern Nigeria, specifically Kontagora in Niger State. The genetic engineering process involves treating the cowpea species with chemical mutagen (ethyl methyl sulphonate or EMS) and physical mutagen (gamma irradiation). Both treated and control (untreated) seeds were planted and observed under water-stressed (dry) conditions for drought-tolerant behaviors in the botanical garden of the Federal University of Technology, Minna, Nigeria.

To guide the metaheuristic search for an optimal cowpea solution with the desired drought-tolerant genetic variability, the mutation-induction process was modeled computationally using a genetic algorithm. The essence of the computational modeling and subsequent implementation using Python programming was to predict when an optimal solution would be obtained as well as show patterns of uncertainties in the stochastic (non-deterministic) optimization problem.

This study aims to show empirically that predictive analytics can be applied in genetic engineering to resolve stochastic optimization problems. The specific objectives are:
i. show that genetic engineering can be used to induce desired traits in crops.
ii. model the genetic engineering of cowpeas using genetic algorithm
iii. demonstrate that the genetic engineering of cowpeas is a stochastic optimization problem

The search and optimization problem in genetic engineering underscores the need to apply predictive analytics like genetic algorithms in this field. For a given species, controlled modification of the chromosome to achieve desired traits takes generations and happens amid uncertainties and therefore needs predictive analytics for accurate bioinformatic information that guides decision-making (E. Okewu, et.al., 2023). The cardinal objective of genetic engineering is to insert random genes in offspring to sustain the genetic diversity in a defined population in a bid to prevent premature convergence. GA has a mutation operator as one of its three operators, alongside a selector operator and crossover operator (S.M. Lim., et.al., 2017) While the selector operator prioritizes individuals with better fitness scores, giving them the latitude to transfer their genes to succeeding generations, the crossover operator represents mating between individuals selected during the selection process. The choice of crossover sites is made randomly and the genes in these sites are exchanged to create completely new offspring or individuals. Any of these operators (or combination) is used by GA to evolve a generation after creating it as the starting generation. Upon specifying a target string, GA progresses to produce the same using a random string with a similar length. The generated string is represented using special symbols and normal characters like A-Z, a-z, and 0-9, which are considered genes. This string is referred to as a solution chromosome or individual. The fitness score is the number of characters in the random string that are different from the target string characters at a particular index. As a result, an individual with a lower fitness score is given priority. A fitness score indicates the ability of an individual to compete. The application of predictive analytics in genetic engineering involves using stochastic search algorithms like GA to search for a solution (chromosome) with optimal or near-optimal fitness value as a convergence criterion. The algorithm keeps track of the population of solutions along with their respective fitness scores to enable individuals with better fitness scores to reproduce themselves faster than others. Similarly, individuals with better fitness scores are selected to mate in a deliberate effort to produce better offspring via the combination of parents' chromosomes. Given the static nature of the population, there is a need to create space for new arrivals. In the process, some individuals die and are replaced by fresh arrivals, giving rise to a new generation at the end of the old population's mating opportunity. As the least-fit solutions die, better solutions emerge over successive generations. Studies have shown that on average, each generation has better genes than the solution of previous generations, implying that successive generations have better solutions than those of past generations. Convergence of the genetic engineering process occurs when an offspring produced is not substantially different from the offspring of previous populations. Predicting convergence using GA is considered as a stochastic optimization because, in every implementation, GA commences with a new random string, resulting in a different output (M. Liu, 2016). Also, predictive analytics algorithms sometimes get stuck at a local optimum solution but can be improved upon using tweak mutation and crossover operators or by updating the calculation algorithm used for the fitness score.

Constrained Stochastic Optimization Problem

Solving constrained stochastic optimization (CSO) problems is hard and also it cannot be solved using normal linear or nonlinear optimization (] M. Mohamed., 2023). The characteristic of CSO problems is that some or all the variables are random variables. In a given system such as the genetic engineering system, the use of random variables is to represent uncertainties in the system. The variables are used in the event there is fluctuation in the parameters of the problem in a given range of values. They are also used when it is hard to assess their expected values. CSO problems can be found in domains like genetic engineering, communication networks, transport engineering, etc. Designers in such environments ents use CSO models in that the systems have to be modeled within the range of mid to long-term period. Predictive analytics and stochastic optimizers like GA are used in solving CSO problems since using conventional methods can be complicated; GA leverages computational techniques to offer accurate and simple solutions (L.A. Sanabria., 2004). The process of mutation in genetic engineering is computationally modeled as a string manipulation in GA and the algorithm progresses from the initial solution (random string) to the optimal solution (target string) using mutation and selection operators. In crop improvement programs, there are biotic (insects, pests, rodents, etc.) and abiotic (sunshine, wind, rainfall, etc.) constraints and uncertainties which are represented by the random string. These constraints and uncertainties can impact genetic diversity. In this regard, genetic mutation or mutation-induced breeding is a classic example of the constrained stochastic optimization problem.

Review of Related Works

The study in (Y. Mohamed Y., et.al., 2020) used genetic engineering for crop improvement in a bid to improve general polygenic features and yield. The research created genetic variation using chemical mutagens to induce a male sterile system in cowpea which facilitates hybridization but is lacking in cowpea. Other objectives achieved included analysis of the impact of chemical mutagens on different morphological features, studying M1 generation genetic diversity, and stipulating LD 50 benchmark for Ethyl Methane Sulphonate (EMS). The mutation-induction in the Vamban 2 cowpea cultivar was executed with the use of eight treatments (10, 20, 30, 40, 50, 60, 70 and 80 mM) of EMS. This was followed by obtaining the LD50 values on account of observations of seed germination, root length, and shoot length under laboratory conditions. The M1 generation was raised under field conditions to assess parameters such as single 100 seed weight, plant yield, number of pods per plant, number of seeds per pod, pollen fertility, number of branches per plant, plant height at maturity, and germination of seeds. The results showed that increased EMS concentration had a negative correlation with phenotypic expression and yield characters. The study observed that the production of cowpeas is an essential component of sustainable agriculture as well and the crop is a good source of protein which contains amino acids like lysine and tryptophan. The article reported that cowpea has free metabolites or other toxins, and it is a nutritious grain legume produced extensively in arid and semi-arid tropics. However, the genetic engineering process adopted by this study used only chemical mutagen while the present study uses chemical mutagen (EMS) and physical mutagen (gamma irradiation) for mutation-induction. Also, the target of the GE approach in this study is the induction of a cowpea male sterile system responsible for hybridization while our GE effort in this present research targets drought tolerance in cowpea.

In (Udhaya KD., 2019), the researchers found out that the combination of gamma rays and EMS in the genetic engineering of moringa plants created mutations with decreased biological damage, leading to the recommendation that both mutagens should be used to induce moringa for desirable qualities. Before arriving at this conclusion, the study induced the PKM-1 variety of moringa with gamma rays and EMS. When compared with the untreated control, it was observed that an increased dose of the mutagens mitigated the probability of seed germination and survival. The study confirmed the fact that the essence of plant breeding using techniques like genetic engineering is to change and better the genetic structure of crops to meet farmers' specific demands. Because natural breeding has the challenge of low genetic variation, breeders are resorting to mutation breeding. In genetic engineering practices, EMS and gamma irradiation (application of gamma rays) are now generally used in inducing mutation in many plant varieties. Though the research focused on using both physical and chemical mutagens to achieve many desired traits in moringa plants, our current work focuses on using both mutagens to induce drought tolerance in cowpeas. Also, the present research formulated genetic mutation as a constrained stochastic optimization problem with the application of predictive analytics (non-exhaustive method) like GA to solve the CSO problem.

The constrained stochastic optimization (CSO) problem is the focus of the authors in (L.A. Sanabria., et., al. 2004) CSO problem is defined as a problem that is difficult to solve or better still, cannot be resolved using linear or nonlinear optimization. In CSO problems, some or all the variables are random variables that are used to qualify uncertainties in a given system. The variables are used to show fluctuation in problem parameters in a specific range of values. They are also used when it is hard to assess their expected values. Such problems are present in genetic engineering, communication networks, transport engineering, etc. In these fields, CSO models are used because the design of the systems is on a mid to long-term basis. Because solving a CSO problem using conventional mathematical methods is complex, the study recommended computational techniques like GA which offers simple and accurate solutions. The study demonstrated the use of GA by finding the optimum design of an Intranet server using the algorithm. The work advocates the use of GA in solving optimization problems, including CSO problems. Our current research aligns with this study as predictive analytics in genetic engineering is viewed as a CSO problem, a claim to be assessed using the mutation-induction of cowpea species from Northern Nigeria. We equally use GA to formulate the mutation process as a string manipulation, followed by the implementation of the algorithm in Python programming.

Liu in (M. Liu ., et., al., 2016) stressed that GA is used in several fields such as bioinformatics, economics, engineering, manufacturing, etc. The authors stressed the utility of stochastic optimization problems in power electronics and control systems for the fact that there is a need to select optimum parameters that offer the least noise effect and maximum control effect in many designs. Resolving such a problem is complex using an exhaustive search method especially when the search domain is huge or infinite. Instead, heuristic search algorithms such as GA is used. There is difficulty associated with evaluating real-life problems with noise and huge computation is needed. As such, the study proposed for such a real-life problem a solution that involves a single objective GA that incorporates computing budget allocation in the selection operator rather than using it during evaluation of fitness. The researchers also studied multi-objective GA which compares the integration of various methods of calculating budget allocation in any of the assessments or the steps in environmental selection. The comparisons take into consideration various levels of noise that are executed on stochastic problems derived from multi-objective optimization problems that are standard. Though the work classified GA as a stochastic optimizer, it did not specify how GA could be used in modeling predictive analytics in genetic engineering in this present research.

In (Q. Cui., et., al. 2020), the authors investigated drought tolerance in Arkansas cowpea species. The study used drought-tolerant lines as parents for breeding and emphasized that the crop is a leguminous crop that exhibits some level of natural drought resistance. As a result, many cowpea lines can survive under hot and dry conditions for upward of 40 days. To confirm this, the researcher used 36 University of Akansa cope lines in screening for drought tolerance at the seedling stage. The experiment took place in a greenhouse using a randomized complete block design (RCBD) to arrange two replicates in a split-plot format. After subjecting the cowpea to drought stress for four weeks, three characteristics that indicate drought tolerance were observed and analyzed. Although we agree with the work that cowpea has some drought-resistant abilities as evident in their capability to survive in environments with scanty rainfall, there is a need to apply genetic engineering to deepen the genetic variability of cowpea to withstand increased climate-induced drought-prone conditions in Northern Nigeria. This explains why the present study is leveraging mutation induction for this purpose.

None of the works formulated predictive analytics in genetic engineering as a constrained stochastic optimization problem, the core consideration of this present research. Also, none modeled mutation-induction for drought tolerance using GA and implemented the same using Python programming to prove experimentally that predictive analytics in genetic engineering is a constrained stochastic optimization problem.

Methodology

Three landraces of cowpea namely; *Dan muzakkari, Gidigiwa, and Dan mesera*, were collected through direct contact with the farmers in Kontagora, Niger State, in Northern Nigeria. Kontagora is known as one of the largest producers of cowpea in Niger State, Nigeria (M.A. Maikasuwa., and A.A. Izo, 2020). The selection of the landraces was based on their yield potential and known drought-tolerant genotypes.
The three landraces of cowpea were subjected to mutagenic treatments using gamma irradiation and Ethyl Methane Sulphonate (EMS). Different doses of gamma irradiation, including 200, 400, 600, and 800 Gray, were applied to the cowpea landraces. The selection of these doses was based on previous studies that have demonstrated their effectiveness in inducing desirable mutations (E.S. Savitri, and S.M. Fauziah., 2020).

Additionally, EMS was used as a mutagenic treatment at a dosage of 0.372% v/v. This particular dosage has been identified as the optimum treatment for inducing mutations in cowpeas (Y. Mohamed Y., et., al., 2020). The mutagenic treatments were done following the guidelines and protocols recommended by the FAO/IAEA Agricultural and Biotechnology Laboratory in Seibersdorf, Austria. The cowpea seeds were soaked with distilled water for 1hr:30mins, after which the water was drained from the soaked seeds, then 1% of ethyl methane sulphonate was prepared to make 100ml with distilled water as the stock solution. 3ml of the stock solution was taken to make up with 97ml of distilled water, making a total of 100ml. The diluted ethyl methane sulphonate was used to soak the seeds again for another 1hr:30mins after which the cowpea seeds were washed with clean water before planting.

The experimental site for this study is located at the experimental garden, Department of Plant Biology, Federal University of Technology, Minna, Nigeria. Minna is geographically located in the North Central Zone of Nigeria, within longitude 6^{o} 33' East and latitude 9^{o} 37' North. It is a grassland savannah area and has a tropical climatic condition with a mean annual temperature, relative humidity, and rainfall of 20-30^{o}C, 61.00%, and 1334.00 cm respectively. The climate brings about two seasons: a rainy season between May and October and a dry season between November and April (F. Odegbenro., 2022).

The experimental design employed in this study is a Complete Randomized Design (CRD). A total of thirty treated seeds including the control were assigned identification numbers and replicated three times to make a total of ninety pots altogether. For each cowpea landrace, six seeds were planted in 10-litre plastic pots that were perforated at the base. The pots were filled with 10 kg of topsoil. After 8 days of planting (DAP), the plants were thinned to four seeds per pot to ensure optimal growth.

Water stress treatment was induced at the vegetative stage; this was imposed after 14 days of initial growth of the plants. Initially, the soil moisture level for all three cowpea landraces was maintained at field capacity, which corresponds to 50% of the maximum water-holding capacity. The well-watered treatment (no stress) was continuously maintained at field capacity throughout the experiment. Soil moisture levels were monitored using a soil moisture meter (MO750, Extech Instruments, USA).

To determine the effects of gamma irradiation and ethyl methane sulphonate (EMS) on cowpeas, selected plants from the different landraces were screened for both qualitative and quantitative traits. The quantitative traits were parameters such as number of pods per plant, pod length, stem diameter, seed yield, and seed weight per plant. These traits were measured using appropriate instruments and techniques, ensuring accurate and consistent data collection.

In addition to the quantitative traits, qualitative parameters related to seed characteristics, such as shapes and colors, were observed and scored. The qualitative scoring provided valuable information on the visual characteristics of the mutant lines and their comparison to the original landraces.

The data obtained on quantitative characters were subjected to analysis of variance (ANOVA) to determine the level of significance among the treatments while the post hoc test was carried out using Duncan's Multiple Range Test (DMRT) to separate the means where necessary using SPSS software version 18.

Computational Experiment

To provide information on progress made in the genetic variability of the mutant lines in each generation, the mutation process was computationally modeled using a genetic algorithm and implemented using Python programming. The genetic algorithm for the predictive analytics is as follows:

```
1) Target String (desired cowpea chromosome/genotype) = mutation-induced drought-tolerant cowpea

2) Length (Target String) = 40

3) Random String (initial cowpea chromosome/genotype) = tttaaa!!!!&&&&&7777%%%%%5555ffff999$$$rr

4) Fitness Score (Genetic Distance) = the number of characters in the Target String that are different from those in the Random String

5) While Fitness Score (Genetic Distance) > 0 repeats:
   a) Choose parents from the population
   b) Implement mutation on a new population to get New String
   c) Calculate the Fitness Score (Genetic Distance) for the new population
   d) Random String =     New String
```

The primary objective of the above algorithm is to conduct a heuristic search for a cowpea solution with optimal drought-tolerance trait. Hence the genetic mutation is modelled as a string manipulation using the target string `mutation-induced drought tolerant cowpea`.

The flowchart for the mutation-induced drought-tolerant trait process is shown in Figure 1.

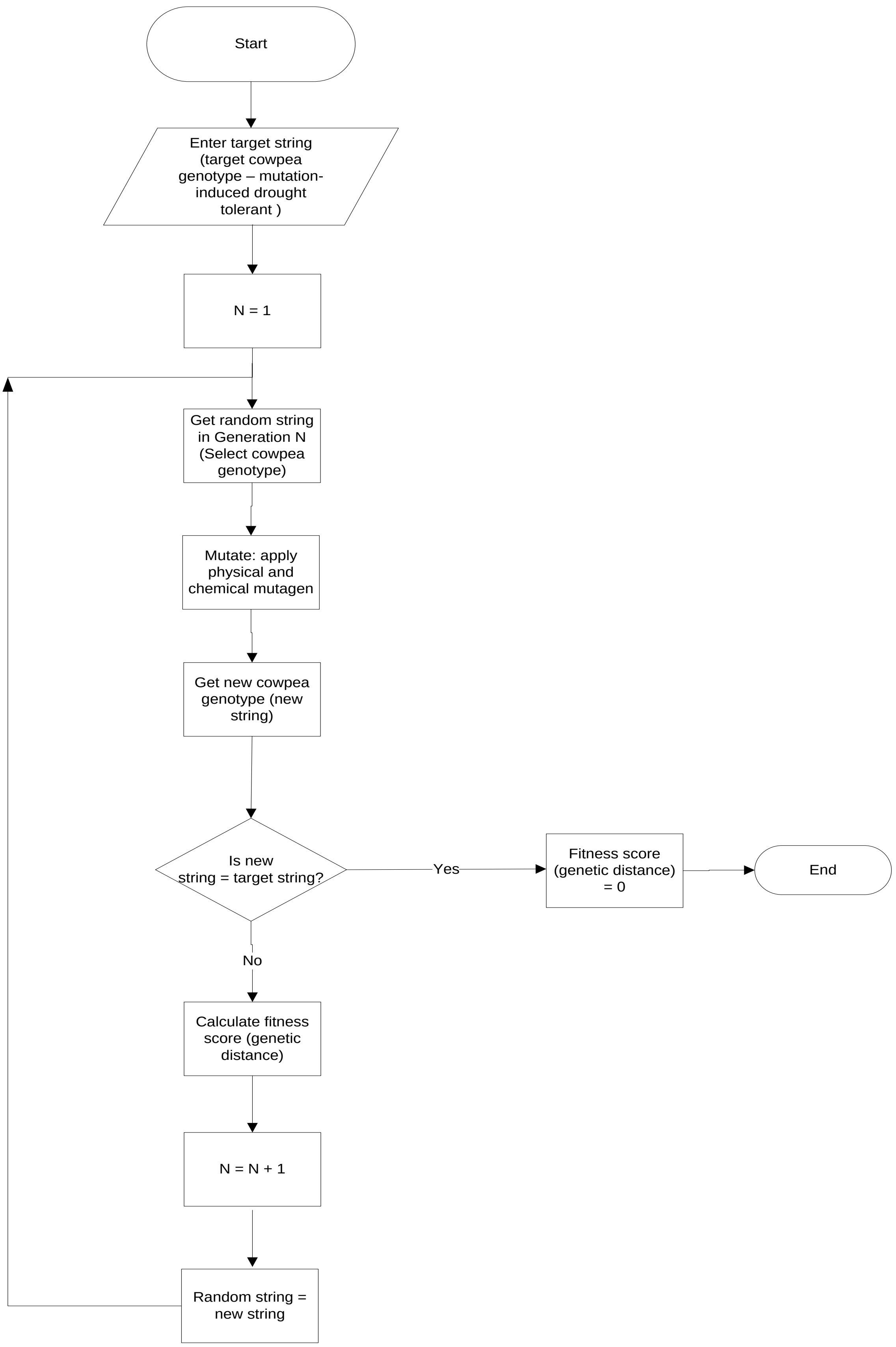

Figure 1. Process flow for predictive analytics in genetic engineering

Results and Discussion

The research aimed to unravel the intricate impact of gamma irradiation and Ethyl Methane Sulphonate (EMS) treatments on agro-morphological and yield parameters of three distinct cowpea varieties namely *Dan mesera* (V1), *Dan muzak kari* (V2) and *Gidigwa* (V3). The results unveiled a diverse array of responses across treatments and durations, contributing to a comprehensive understanding of the influence of these treatments on cowpea growth.

Effects of Gamma Irradiation and Ethyl Methane Sulphonate on Agro-Morphological Parameters (Plant Heights).

Table 1 shows the results of a study on the effects of gamma irradiation, ethyl methane sulfonate (EMS), and their combination on plant height (PH) of three varieties of cowpea (V1, V2, and V3). The study was conducted at different durations after planting (2 weeks, 4 weeks, 6 weeks, and 8 weeks respectively).

Table 1. Effects of Gamma Irradiation and Ethyl Methane Sulphonate on Plant Heights

Parameter	PH2	PH4	PH6	PH8
V1	19.50±1.32cde	27.30±3.74g	30.87±3.01ef	33.07±3.45a
V2	22.90±0.36e	23.80±0.17cdefg	31.90±3.01f	33.10±2.69a
V3	19.90±1.55cde	24.67±1.33defg	30.77±3.12ef	32.20±3.65a
V1G1	14.34±2.07abc	15.87±2.03ab	19.63±2.71ab	23.37±1.39a
V1G2	18.67±2.46bcde	22.60±1.16bcdefg	26.37±1.59bcdef	27.40±1.68a
V1G3	18.90±2.08bcde	22.77±1.41bcdefg	27.37±3.06bcdef	29.20±2.72a
V1G4	20.27±1.37cde	25.50±1.10efg	28.60±2.77cdef	30.10±3.10a
V1E	19.47±1.89cde	23.43±0.70defg	24.57±0.43abcdef	27.93±1.57a
V2G1	16.73±0.72bcd	21.57±0.73abcdefg	25.10±1.45abcdef	26.57±1.22a
V2G2	18.10±0.95bcde	22.90±1.32bcdefg	25.17±1.48abcdef	27.07±1.48a
V2G3	17.40±3.32bcde	20.97±2.77abcdef	22.40±2.55abcde	25.50±3.79a
V2G4	17.67±0.27bcde	22.50±1.54bcdefg	24.33±1.87abcdef	25.67±1.84a
V2E	15.73±0.50bcd	21.67±0.29abcdefg	27.23±1.58bcdef	28.10±1.50a
V3G1	19.43±2.75cde	23.33±3.00cdefg	28.50±2.08cdef	30.80±1.78a
V3G2	15.20±0.20bcd	19.93±0.87abcde	24.60±2.71abcdef	26.83±2.20a
V3G3	18.77±1.55bcde	21.87±1.30abcdefg	26.73±3.32bcdef	28.73±3.41a
V3G4	17.00±1.01bcde	20.00±1.04abcde	22.17±1.70abcd	26.70±0.79a
V3E	19.40±3.05cde	22.20±2.29bcdefg	25.27±0.87abcdef	28.50±0.90a

V1G1E	17.80±2.76bcde	23.97±1.33defg	28.37±2.66cdef	31.63±1.73a
V1G2E	18.46±1.43bcde	22.83±3.57bcdefg	25.07±2.60abcdef	26.80±2.31a
V1G3E	20.87±2.51de	26.37±0.98fg	29.07±2.15cdef	32.00±2.31a
V1G4E	14.87±1.93abcd	18.23±3.49abcd	21.37±4.48abc	21.50±3.96a
V2G1E	19.47±0.79cde	21.47±0.58abcdefg	30.33±5.36def	30.23±5.64a
V2G2E	13.07±0.57ab	17.10±0.40abc	19.10±0.71ab	22.27±2.38a
V2G3E	9.57±1.55a	15.00±1.45a	17.07±0.91a	19.67±0.96a
V2G4E	16.47±0.47bcd	20.23±0.23abcdef	23.10±0.89abcde	28.30±3.55a
V3G1E	19.00±3.06bcde	22.17±2.03bcdefg	26.27±1.60bcdef	27.47±1.78a
V3G2E	16.50±0.90bcd	21.97±0.98abcdefg	26.50±2.25bcdef	28.07±2.72a
V3G3E	16.67±1.36bcd	22.17±2.49bcdefg	24.43±1.84abcdef	26.33±1.20a
V3G4E	18.17±0.73bcde	22.90±1.19bcdefg	25.30±0.70abcdef	112.23±5.89b

Values are Mean±Standard Error of Mean. Means with the same letter(s) within a set of treatment column are not significantly different at $p \leq 0.05$ using Duncan Test

The results (Table 1) showed that the PH of all three varieties of cowpea increased with duration after planting. However, the PH of the varieties treated with gamma irradiation and EMS was significantly lower than the PH of the varieties that were not treated.

For example, at 2 weeks after planting, the PH of V2 that was not treated with either gamma irradiation or EMS recorded the highest value of 22.90cm while the same variety2 treated with a combination of gamma and EMS recorded the lowest value of 9.57cm which shows that both gamma and EMS has negative effects on plant height at week2.

The results further revealed that the combination of gamma irradiation and EMS had a more negative effect on PH. For example, at week 4 and week 6 after planting, the PH of both variety1 and variety2 control treatments recorded 27.30cm and 31.90cm respectively which were the highest while the combination of gamma and EMS treatments showed the lowest values of 15.00cm and 17.07cm at week 4 and 6 respectively.

Effects of Gamma Irradiation and Ethyl Methane Sulphonate on Yield Parameters

Table 2 shows the results of a study on the effects of gamma irradiation and ethyl methane sulfonate (EMS) on the number of pods, length of pods, number of seeds per pod, weight of seed pod, and weight of 100 seeds of three varieties of cowpea (V1, V2, and V3).

Table 2 Effects of Gamma Irradiation and Ethyl Methane Sulphonate on Yield Parameters.

Parameter	No of Pods	Length of Pods	No of Seeds/Pod	Weight of of Seeds/Pod	Weight of 100 Seeds
V1	35.00±6.66ef	11.00±0.70abc	8.20±0.88abcd	2.09±0.12ab	23.60±0.42c
V2	35.33±3.71f	15.00±0.00ghi	11.70±0.76efgh	3.11±0.20cdef	16.80±0.70ab
V3	26.33±9.02bcdef	14.50±0.50fghi	10.10±0.94bcdefgh	3.16±0.40cdef	23.50±0.00c
V1G1	10.00±1.15abc	12.60±0.65cde	10.00±0.77bcdefgh	3.11±0.26cdef	24.10±0.50cd
V1G2	28.67±7.19cdef	11.30±0.75abc	9.40±0.90abcdef	2.43±0.32abcdef	22.01±0.65c
V1G3	10.67±4.26abc	13.00±0.45def	11.00±1.13defgh	2.69±0.20bcdef	23.00±0.75c
V1G4	13.67±5.21abcd	10.10±0.64a	7.60±0.64ab	2.16±0.21abcd	24.22±0.37cd
V1E	14.00±5.03abcd	13.00±0.47def	10.00±0.00bcdefgh	3.00±0.26cdef	24.63±0.64cd
V2G1	32.33±8.41def	13.70±0.42defgh	10.70±0.92defgh	3.45±0.24ef	25.20±0.47cd
V2G2	16.00±2.08abcdef	13.80±0.25efghi	9.90±0.71bcdefgh	2.96±0.25cdef	25.50±0.51cd
V2G3	15.33±2.40abcdef	13.30±0.52defg	9.80±0.95abcdefgh	2.56±0.31bcdef	22.30±0.71c
V2G4	16.33±6.84abcdef	13.50±0.31defg	8.60±0.73abcde	2.69±0.23bcdef	27.00±0.36de
V2E	14.67±4.81abcdef	14.50±0.34fghi	11.10±0.91defgh	3.61±0.21def	33.30±0.33e
V3G1	14.33±6.12abcde	15.00±0.45ghi	11.30±0.83efgh	3.03±0.23cdef	19.10±0.38b
V3G2	17.67±4.10abcdef	13.60±0.48defg	12.20±0.59efgh	2.98±0.20cdef	19.90±0.45b
V3G3	12.00±3.61abcd	15.50±0.37i	12.60±0.52h	3.36±0.25def	13.00±0.48a
V3G4	13.33±2.33abcd	14.50±0.54fghi	12.40±0.54efgh	3.25±0.20def	22.20±0.37c
V3E	18.00±4.62abcdef	15.40±0.40hi	10.10±1.22bcdefgh	2.34±0.17abcde	17.60±0.54ab
V1G1E	6.67±3.18ab	13.10±0.38def	10.00±0.84bcdefgh	3.02±0.20cdef	24.30±0.40cd
V1G2E	17.33±1.48abcdef	12.60±0.69cde	8.30±1.01abcd	2.46±0.31abcdef	23.20±0.38c
V1G3E	24.67±7.80abcdef	10.80±0.33ab	6.90±0.77a	1.82±0.22a	21.60±0.69c
V1G4E	4.00±2.00a	11.30±0.79abc	9.80±0.70abcdefgh	2.86±0.21cdef	23.40±0.33c
V2G1E	15.00±2.89abcd	13.30±0.87defg	10.30±0.82bcdefgh	3.01±0.35cdef	26.70±0.79cd

V2G2E	12.33±3.18abcd	12.00±0.65bcd	8.20±1.54abcd	2.59±0.44bcdef	23.50±0.87c
V2G3E	8.00±1.15abc	14.60±0.31fghi	9.60±0.88abcdefgh	2.88±0.25cdef	24.60±0.65cd
V2G4E	21.33±6.64abcdef	13.40±0.65defg	10.60±0.69cdefgh	3.71±0.17f	22.40±0.31c
V3G1E	20.00±6.56abcdef	13.80±0.51efghi	9.50±0.96abdefg	2.46±0.19abcdef	19.50±0.45bc
V3G2E	18.67±4.06abcdef	13.90±0.41efghi	11.90±0.85efgh	2.54±0.15bcdef	17.00±0.36ab
V3G3E	21.33±6.33abcdef	14.90±0.31ghi	12.50±0.85gh	3.13±0.16cdef	19.00±0.32b
V3G4E	13.33±1.67abc	14.40±0.31fghi	7.70±1.02abc	2.15±0.32abcd	21.30±0.53c

Values are Mean±Standard Error of Mean. Means with the same letter(s) within a set of treatment columns are not significantly different at *$p \leq 0.05$* using Duncan Tests

Aa – no significant difference
Ab – there is a significant difference
Abc – a higher significant difference

The table shows the number of pods (No of Pods), length of pods (Length of Pods), number of seeds per pod (No of Seeds Pod), weight of seed pod (Weight of Seed Pod), and weight of 100 seeds (Weight of 100 Seeds) of all three varieties of cowpea were affected by gamma irradiation and EMS.

Predictive Analytics in the Genetic Engineering Process
Tables 3, 4, and 5 show the results of studies on the computational modeling of the mutation-induced drought tolerance of the local cowpea species (*Dan muzakkari, Gidigiwa, and Dan mesera*). The aim is to forecast bioinformatic outcomes for informed decision-making in investigation and breeding efforts. Earlier studies has shown that genetic algorithms can be used to forecast bioinformatic outcomes of mutation-induction [7].

Each of the three experiments has the target string (optimal solution) of "mutation-induced drought-tolerant cowpea" derived at a fitness score (genetic distance) of 0 while each experiment generated a different initial random string. The random string is a computational representation of the uncertainties in the genetic engineering process imposed by biotic factors (insects, pests, rodents, micro-organisms, etc) and abiotic factors (sunshine, wind, rain, etc). The ability of GA to generate random strings and also generate different random strings for each experiment despite solving the same problem is an affirmation that GA is a constrained stochastic optimizer [5], [9].

The results of Experiment 1 in Table 3 show that the algorithm converged at 9412th with a fitness score (genetic distance) of 0. In essence, the algorithm is predicting that the optimal cowpea solution with the desired drought-tolerant feature would be obtained at the 9412th generation represented by the string (chromosome) mutation-induced drought-tolerant cowpea.

Table 3. Experiment 1 results

Generation	Random String	Fitness Score
Generation 1	`Jdt?iiG&5"Bq=g0dD/B4P9evy[- 4gunLBkK&mw)`	36
Generation 2	`String: Jdt?iiG&5"Bq=g0dD/B4P9evy[- 4gunLBkK&mw)`	36
Generation 3	`Jdt?iiG&5"Bq=g0dD/B4P9evy[- 4gunLBkK&mw)`	36
Generation 4	`Jdt,iiG&"Bq,g0d;C{qPqevy[! 4gun2BKKw;w)`	35
Generation 5	`2B(PldT/WwnFjc dgMRn tyHYd)2e}2{j"TKwIea`	33
Generation 6	`2B(PldT/WwnFjc dgMRn tyHYd)2e}2{j"TKwIea`	33
Generation 7	`J3t8Mion3"n1,budpZUPqTHXtL e&Z;2cX}2}e)`	31
Generation 8	`J3t8Mion3"n1,budpZUPqTHXtL e&Z;2cX}2}e)`	31
Generation 9	`J3t8Mion3"n1,budpZUPqTHXtL e&Z;2cX}2}e)`	31
Generation 10	`J3t8Mion3"n1,budpZUPqTHXtL e&Z;2cX}2}e)`	31
….	….	…
Generation 94 09	`mutation-induced dr}ught tolerant cowpea`	1
Generation 94 10	`mutation-induced dr}ught tolerant cowpea`	1
Generation 94 11	`mutation-induced dr}ught tolerant cowpea`	1
Generation 94 12	`mutation-induced drought-tolerant cowpea`	0

Table 3 shows further that in the first generation (M1), the fitness score (genetic distance) is 36 while the random string is `Jdt?iiG&5"Bq=g0dD/B4P9evy[- 4gunLBkK&mw`.

In Table 4, outcomes of Experiment 2 are shown. The GA converged at the 9717th generation with fitness score (genetic distance) of 0.

Table 4. Experiment 2 results

Generation	Random String	Fitness Score
Generation 1	`f%evEqv)-%s1;)J#&HrJ4UgwgKF& w?N/jqjB#"`	38
Generation 2	`f%evEqv)-%s1;)J#&HrJ4UgwgKF& w?N/jqjB#"`	38
Generation 3	`Bu;R]1ncWia8j2]V18XHZgET:]F?v:HdlW.t#Uqa`	36
Generation 4	`Bu;R]1ncWia8j2]V18XHZgET:]F?v:HdlW.t#Uqa`	36
Generation 5	`iVt!(6ULGyyKuidFh1o5]JtsQDqT09A7 kM/qT`	35
Generation 6	`i1tu;onIZ}-foiR$YGx&gtDYsf 0H9MG Ho/Y[4`	34
Generation 7	`dus![on]-iqX6#]g KrYQgTtop4?]/K]z0jwcp6D`	32

Generation 8	dus![on]-iqX6#]g KrYQgTtop4?]/K]z0jwcp6D	32
Generation 9	Cut(uto]!A&u%O_V QJl&ghR KfS=H:K5 HS!p{	31
Generation 10	Gut(u;opJZ-#{Oi71oJLZght ?f M2a$5 5oYp5z	29
….	….	…
Generation 9714	mutation-ind=ced drought tolerant cowpea	1
Generation 9715	mutation-ind=ced drought tolerant cowpea	1
Generation 9716	mutation-ind=ced drought tolerant cowpea	1
Generation 9717	mutation-induced drought tolerant cowpea	0

However, in the same Experiment 2 (Table 4), the random string in M1 is f%evEqv)-%s1;)J#&HrJ4UgwgKF& w?N/jqjB#" though the fitness score (genetic distance) is 38.

In Table 5, the results obtained during Experiment 3 are presented. The convergence of the algorithm, which indicates the end of the metaheuristic search for an optimal cowpea solution, happened at the 14338th generation.

Table 5. Experiment 3 results

Generation	Random String	Fitness Score
Generation 1	qSE==hTm/2K8c1d$J:8ANAJvs(p=%qw ;;yv/g	37
Generation 2	qSE==hT m/2K8c1d$J:8ANAJvs(p=%qw ;;yv/g	37
Generation 3	dS?m]IL/g//cwH$t:dBgpA?aXp4&G}X Vw/Y=@	36
Generation 4	dS?m]IL/g//cwH$t:dBgpA?aXp4&G}X Vw/Y=@	36
Generation 5	$fEZVim6]iDK8A?d1JHTxNALtwp=Xqw) nn#v@H	35
Generation 6	Uf#ki36wioD],ed MyPh5$%JmmD_-C39 2[GY7?	34
Generation 7	Tne SiY.mID/GQed$dy9tg]LJq$-4vz9 2oC2=S	33
Generation 8	{{,Zvi.F&lGI4Q?dIdy-8gtwQtd-Q0G)9 c75p%m	32
Generation 9	qu?[=i nwiiLMc?Hd,8&NW1?1HueFq)t Yoy2z3	30
Generation 10	TuE[=i nwI4Lsc5AI8&NoL?tHueFz)t coY2pO	29
….	….	…
Generation 14335	mutation-induced drought toEerant cowpea	1
Generation 14336	mutation-induced drought toEerant cowpea	1
Generation 14337	mutation-induced drought toEerant cowpea	1
Generation 14338	mutation-induced drought tolerant cowpea	0

Meanwhile, Experiment 3 (Table 5) produced a random string with the fitness score (genetic distance) of 37 at in the first generation (M1). The random string is `qSE==hTm/2K8c1d$J:8ANAJvs(p=%qw;;yv/g`.

The results of the three experiments indicate that the algorithm sometimes gets stuck at a local optimum solution (plateau). In Experiment 1, for example, this happened in generations 5 and 6. In Experiment 2, generations 7 and 8 experienced a plateau while in Experiment 3, the algorithm got stuck at the local optimum in generations 14335, 14336, and 14338. However, improving on the plateau is achievable by either tweaking the mutator operator or updating the fitness score calculation algorithm.

The above empirical outcomes have confirmed that applying predictive analytics in genetic engineering is a constrained stochastic optimization problem. It also shows that GA exhibits stochastic optimization like deep neural network (DNN) (E. Okewu., et., al., 2022). This is evident in the fact that in the same problem of applying predictive analytics in the genetic engineering of cowpeas for drought-tolerant traits to withstand climate-induced drought-prone conditions in Northern Nigeria, GA generated different random strings in each of the three experiments. Also, its convergence in the three experiments happened at different generations. While both GA and DNN can be used to solve CSO problems, their modus operandi differ. DNN uses random variables while GA uses random string in their heuristic search and optimization aimed at converging to an expected value. While the difference between the actual value and expected value in GA is referred to as fitness score, in DNN it is called error value. Both stochastic optimizers converge at the point where the fitness score/error is within the limit of tolerance or better still, zero.

From the aforementioned, the three (3) specific objectives of this work have been achieved: the analysis of the agro-morphological and yield parameters showed that genetic engineering of the cowpea cultivars indeed induced drought-tolerant traits. Also, the genetic engineering process was modeled using a genetic algorithm implemented using Python programming. Outcomes of the computational experiments (execution of the Python programs) that the genetic engineering of cowpea is a stochastic optimization problem

Conclusion

Genetic engineering and gene drives are important in the conservation of crops (R. Sandler., 2020). The search for native cowpea species that can withstand the increased climate-induced drought-prone conditions in Northern Nigeria, and by extension its conservation, is the motivation for this work. For indigenous cowpea cultivars to survive the climate-imposed drought conditions, their genetic diversity has to incorporate drought-resilient traits. Achieving this using natural breeding is near-impossible. Hence, genetic engineering was applied to the treatment of selected cowpea landraces Northern Nigeria (specifically, from Kontagora in Nigeria State) with chemical mutagen (EMS) and physical mutagen (gamma rays). To guide decision-making by investigators and breeders in the crop improvement program, predictive analytics was applied using genetic algorithm. Results of both field experiments and computational experiments indicate that the cowpea species are developing drought-tolerant traits from generation to generations. While the agro-morphological and yield parameters show positive results, the computational results predict when an optimal cowpea solution would be obtained. The uncertainties and constraints posed by biotic and abiotic factors underscore the fact that predictive analytics in genetic engineering is a constrained stochastic optimization problem.

Conflict of Interest
There is no conflict of interest between the authors in the execution of this study.

References

Antoniou, M. (2021). What is GM and what are the health risks? *ENVI Committee - European Parliament*. https://www.europarl.europa.eu/cmsdata/237985/Antoniou.pdf

Cui, Q., Xiong, H., Yufeng, Y., Eaton, S., Imamura, S., Santamaria, J., Ravelombola, W., Mason, R. E., Wood, L., Mozzoni, L. A., & Shi, A. (2020). Evaluation of Drought Tolerance in Arkansas Cowpea Lines at Seedling Stage. *HortScience*, *55*(7), 1132–1143. https://doi.org/10.21273/hortsci15036-20

L. A. Sanabria, B. Soh, T. S. Dillon and L. Chang (2003) Genetic algorithms in stochastic optimisation, The 2003 Congress on Evolutionary Computation, 2003. CEC '03., Canberra, ACT, Australia, 2003, pp. 815-822 Vol.2, doi: https://doi.org/10.1109/CEC.2003.1299751

Lim, S. M., Sultan, A. B. M., Sulaiman, M. N., Mustapha, A., & Leong, K. Y. (2017). Crossover and Mutation Operators of Genetic Algorithms. *International Journal of Machine Learning and Computing*, *7*(1), 9–12. https://doi.org/10.18178/ijmlc.2017.7.1.611

Liu, M. (2016). *Genetic Algorithms in Stochastic Optimization and Applications in Power Electronics*. https://doi.org/10.13023/etd.2016.475

Maikasuwa, M. A., & Izo, A. A. (2022). Impact of Purdue Improved Cowpea Storage (PICS) Bag on the Profitability of Cowpea Storage in Kontagora Local Government Area of Niger State. *Equity Journal of Science and Technology*, *8*(1), 79–83. https://doi.org/10.4314/equijost.v8i1.13

Mnif, M., & Pham, H. (2001). Stochastic optimization under constraints. *Stochastic Processes and Their Applications*, *93*(1), 149–180. https://doi.org/10.1016/s0304-4149(00)00089-2

Mohamed Yaseen SK, Sravan Naga Parimala Kumar K, Prabakaran S and Suresh S. (2020) Ethyl methane sulphonate (Ems) induced mutations in M_1 generation of cowpea (*Vigna unguiculata* (L.) Walp). Journal of Pharmacognosy and Phytochemistry, 9(2), pp. 1545-1547.

Muthuvel, J., Saharia, M., Kumar, S., Abiala, M. A., Rao, G. J. N., & Sahoo, L. (2020). Progress in Genetic Engineering of Cowpea for Insect Pest and Virus Resistance. In *Springer eBooks* (pp. 115–137). https://doi.org/10.1007/978-981-15-5897-9_7

Odegbenro, F.J. and Ojoye, S. (2017). Assessment of Public Perception of Climate Change Issues in Minna, Niger State, Nigeria. Environmental Technology and Science Journal. 8(1): 14-22

Okewu, E., Adewole, P., & Sennaike, O. (2019). Experimental Comparison of Stochastic Optimizers in Deep Learning. In *Lecture notes in computer science*. pp. 704–715. https://doi.org/10.1007/978-3-030-24308-1_55

Okewu, E., Okewu, K., Maidin, S. S., & Shing, W. L. (2023). Genetic Mutation of Cowpea as a Constrained Stochastic Optimization Problem in Sustainability. *JOURNAL OF DATA SCIENCE* 2023 (13). pp. 1-16. http://eprints.intimal.edu.my/1816/1/jods2023_13.pdf

Okewu, K., Okewu, E., Shing, W. L., & Maidin, S. S. (2023). Genetic Algorithm for Forecasting Bioinformatic Outcomes of Mutation-induced Cowpeas for Sustainable Development. *JOURNAL OF DATA SCIENCE* 2023(12). pp 1-13. http://eprints.intimal.edu.my/1062/1/jods2023_12.pdf

Sandler R. (2020). The ethics of genetic engineering and gene drives in conservation. *Conservation biology: the journal of the Society for Conservation Biology*, *34*(2), 378–385. https://doi.org/10.1111/cobi.13407

Savitri, E. S., & Fauziah, S. M. (2020). ROLE OF ISSR MOLECULAR MARKER IN ANALYSIS OF GENETIC DIVERSITY OF MUTANT Glycine soja THROUGH MUTATION INDUCTION WITH COMBINATION OF GAMMA RAY IRRADIATION AND EMS (ETHYL METHANE SULFONATE). *Journal of Applied Biological Sciences*, *14*(1), 113–123.

Soviadan, Mawussi. (2019). CONSTRAINED AND UNCONSTRAINED OPTIMIZATION. https://doi.org/10.13140/RG.2.2.30215.96163/2

Udhaya Kumar D, Paramaguru P, Swaminathan V, Manikanda Boopathi N, Juliet Hepziba S, Arumugam T and Susmitha D (2019). Effect of gamma irradiation and ethyl methane sulphonate in annual moringa (*Moringa oleifera* L.) variety PKM-1. Journal of Pharmacognosy Phytochemistry 8(5), pp. 2258-2261

Yali, W. (2022). Genetic engineering its application, importance and future aspects in modern Crop improvement. *African Journal of Food Science and Technology* pp. 01–06. https://doi.org/10.14303//ajfst.2022.034