NSGA-II Frame-worked Multi-Objective Evolutionary Optimization for Egg-Sperm Compatibility Simulation

Emirhan Kara [1]

***Abstract*— This project presents a multi-objective genetic algorithm designed to simulate and optimize egg-sperm compatibility. The main goal is to evolve a population of individuals (sperm cells) that maximize both biological quality and genetic compatibility to a target egg cell within a limited resource which is split between genetic and biological traits.**

Two objective functions are defined: genetic compatibility and biological quality. These objectives are computed using a set of biologically inspired decision variables that models the individuals in the population. The decision variables are: HLA profile, DNA fragmentation index (DFI), motility, morphology, velocity, and pH tolerance. The HLA profile and pH tolerance affect the genetic matching, while the remaining traits affect the quality of the sperm. Each variable is influenced by a limited resource budget, which mimics real-world trade-offs between genetic and physical quality.

The genetic algorithm used in the project follows the Elitis Non-Dominates Sorting Algortihm (NSGA-II) framework. For the evaluation step, the crowding distance method is implemented to maintain diversity within the population. In addition to that, a dynamic penalty function is introduced to penalize individuals that cannot satisfy a genetic compatibility threshold. The mating pool is created by using tournament selection method which prioritizes the feasible, less-crowded, and high-ranked individuals. The variation operator, which includes Simulated Binary Crossover (SBX) crossover with modified random mutation creates off-springs from the mating pool. To provide elitism, the survivor stage merges the off-springs with the current population and re-evaluates before creating the next generation using rank-based filtering approach.

This project can be extended to solve broader real-world problems such as *in vitro fertilization (IVF)*, *organ matching*, *personalized medicine*.

Keywords : multi-objective optimization, NSGA-II, egg-sperm compatibility simulation.

I. Introduction

Fertilization is a fundamental biological process that initiates the development of a new organism through the successful interaction of sperm and egg cells. The outcome of this multi-step process is determined by a complex interaction of molecular, cellular, and immuno-genetic mechanisms, ultimately determining whether the gametes are compatible or not. [1], [2]. Studies have highlighted the importance of genetic matching and selective chemical communication in sperm-egg recognition, indicating that environmental parameters like pH and compatibility at the human leukocyte antigen (HLA) level may affect the success of fertilization [3], [4]. In this project, a multi-objective genetic algorithm is developed to simulate and optimize egg-sperm compatibility, focusing on both biological quality and genetic matching.

Genetic algorithms (GAs) are widely used optimization tools inspired by the process of natural evolution. They are especially powerful when solving complex problems with multiple objectives and constraints. The algorithm framework in this project is based on Elitis Non-Dominates Sorting Algortihm (NSGA-II) [5]. NSGA-II introduces the concepts of rank-based sorting and crowding distance to maintain diversity, uses crowded tournament selection to select the mating pool, and uses rank-based filtering to populate the next generation by using the whole population (off-springs and the current population) to provide elitism over generations. In this study, two objectives are defined: genetic compatibility and biological quality. These objectives conflict due to limited available resources, causing an individual (sperm cell) to be either more compatible with the egg cell and less individual quality or the reverse. This trade-off is handled by modeling a fixed resource budget that must be split between genetic and biological traits for each individual in the population. Also a dynamic penalty function is implemented. This function penalizes individuals that do not meet a minimum genetic compatibility threshold to obtain biologically feasible individuals

This study presents the representative presentation of the problem, implementation of NSGA-II based multiobjective genetic algorithm, and comparative results for the different parameter combinations. As a result, it is shown that the algorithm is successfully overcomes the task and the statistical analysis is discussed.

II. Methodology

This section describes the methodology followed in this project including the decision variables creation, objective functions, formulation, and constraints.

A. Decision Variables

To start modeling the problem, the first step is to initialize the attributes of an egg cell. After that, decision variables that identifies each sperm cell in the population is declared. These variables are carefully selected for a simplified yet representative abstraction of the problem [6] [7].

Attributes of an egg cell as the following:

- **HLA profile** – a fixed set of alleles representing the egg's immuno-genetic identity, used for compatibility matching with sperm [6]. The HLA profile consists of six alleles randomly selected from a predefined pool of thirteen distinct alleles.

- **Ideal pH range** – a range in which the sperm's pH tolerance must fall to ensure potential fertilization. It is set to 7.2 - 8.0 [7].

Attributes of an sperm cell (decision variables of the GA) as the following:

- **HLA profile** – affects compatibility with the egg [3] [6]. Length of the HLA profile for a cell is bounded to the range [2, 8]
- **pH tolerance** – determines the sperm's ability to survive in the egg's environment [7]. It is bounded to the range [6.5, 8.5]
- **DNA fragmentation index (DFI)** – reflects the integrity of the genetic material [8]. It is bounded to the range [0, 100]
- **Motility** – indicates the ability of sperm to move towards the egg [9]. It is bounded to the range [1, 100]
- **Morphology** – assesses the size and shape normality of the sperm cell [10]. It is bounded to the range [1, 100]
- **Velocity** – measures the speed of movement [11]. It is bounded to the range [3, 100]

B. Objectives

The first objective function, **genetic compatibility**, is derived by the similarity between the sperm's HLA profile and the egg's HLA profile, as well as their pH compatibility. These factors are correlated with successful fertilization and immunological acceptance[12]. To calculate the second objective function, **biological quality**, the remaining attributes including motility, morphology, DFI, and velocity are used which reflect the functional viability of the sperm [13].

The decision variables used in this model show different correlations with the objective functions. **Genetic compatibility** is positively influenced by the similarity in the HLA profile between the egg and the sperm. The first objective function is also affected by the size of the overlap between pH ranges of the egg and the sperm. On the other hand, higher values of sperm motility, velocity, and morphology, which indicate the sperm's functional capacity to reach and fertilize the egg, have a beneficial impact on the second objective function, **biological quality**. Additionally, DFI value has a negative correlation with biological quality, as it indicates the integrity of the genetic material.

In the simulation, each sperm individual is bound to a fixed total resource budget that needs to be divided between two goals. **Genetic resources** and **biological resources** are two internal variables that control this allocation; they always add up to 100, with a lower bound of 10 and an upper bound of 90. An increased weight in **genetic resources** improves the sperm's possibility of compatibility with the egg. However, as the weight shifts towards the **biological resources**; sperm's motility, morphology, velocity, and DNA integrity are increased. Compatibility and individual quality are fundamentally traded off in this configuration.

C. Formulation

This section presents the mathematical formulation of the two objective functions. The goal is to simultaneously maximize *genetic compatibility* and *biological quality* of sperm individuals relative to a given egg.

1) Symbol Definition:

- h – HLA match score between sperm and egg.
- p – pH compatibility score.
- r_g – proportion of resources allocated to genetic features.
- r_b – proportion of resources allocated to biological features.
- m – motility value of the sperm.
- q – morphology value of the sperm.
- v – velocity of the sperm.
- d – DNA fragmentation index (DFI).
- $\text{norm}(x)$ – normalized value of variable x.

2) Objective 1: Genetic Compatibility: The first objective measures the compatibility of sperm with the egg based on immunogenetic and environmental pH matching.

$$f_1 = (0.7 \cdot h + 0.3 \cdot p) \cdot (0.3 + 0.7 \cdot r_g) \tag{1}$$

where

$$h = \frac{|\text{HLA}_{\text{sperm}} \cap \text{HLA}_{\text{egg}}|}{|\text{HLA}_{\text{egg}}|} \tag{2}$$

$$p = \begin{cases} 1.0, & \text{if } \text{pH}_{\text{tol}} \in [\text{pH}_{\min},\ \text{pH}_{\max}] \\ 1 - \min(|\text{pH}_{\text{tol}} - \text{pH}_{\min}|, & \\ \qquad |\text{pH}_{\text{tol}} - \text{pH}_{\max}|), & \text{otherwise} \end{cases} \tag{3}$$

3) Objective 2: Biological Quality: The second objective captures the physical and structural integrity of the sperm cell.

$$\begin{aligned} f_2 = &(0.3 \cdot \text{norm}(m) + 0.25 \cdot \text{norm}(q) \\ &+ 0.15 \cdot \text{norm}(v) + 0.3 \cdot (1 - \text{norm}(d))) \cdot (0.2 + 0.8 \cdot r_b) \end{aligned} \tag{4}$$

$$\text{norm}(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

with the bounds:

$$\begin{aligned} m &\in [30, 100], \\ q &\in [10, 100], \\ v &\in [10, 100], \\ d &\in [0, 50]. \end{aligned}$$

III. Genetic Algorithm

The genetic algorithm framework of the project is influenced by the NSGA-II algorithm [5]. This section describes each operation of the genetic algorithm including evaluation, selection, variation, and survival.

A. Evaluation

The evaluation stage of the implemented multi-objective genetic algorithm starts by calculating the objective functions as mentioned in Section II-C. After that, to calculate crowding distances, we first rank the population with non-dominated sorting algorithm. The non-dominated sorting algorithm finds pareto-optimal solutions within the population, extracts them from the population and recursively finds the next pareto-optimal solutions until there is no more individuals in the population. After this operation, the population is divided into different-leveled fronts, each containing non-dominated set of solutions within themselves. Ranks for each individual within the same fronts are assigned starting from one being the best and iteratively increasing by one. After the ranks are determined for each individual, the crowding distances are calculated as follows: First, determines the global maximum and minimum values for each objective to be used for normalization. Then, for each objective, individuals in the front are sorted in ascending order according to their value in that objective. The smallest and largest valued individuals are assigned an infinite crowding distance to preserved boundary solution. For the remaining individuals, the crowding distance is calculated using a normalized difference formula given in Equation 6 where (i) represents the individual and (m) represents the m-th objective. The whole process used to calculate ranks and crowding distances is summarized in Algorithm 1

$$\mathrm{CD}(i) \mathrel{+}= \frac{f_{i+1}^{(m)} - f_{i-1}^{(m)}}{f_{\max}^{(m)} - f_{\min}^{(m)}} \tag{6}$$

Algorithm 1 Evaluation

Input: Population P
Output: Fronts F, with *rank* and *crowding distance* assigned
`EvaluateObjectives`(P)
$F \leftarrow$ `NonDominatedSort`(P)
for each front $F_k \in F$ **do**
  `AssignRank`(F_k)
  `AssignCrowdingDistances`(F_k)
end for

B. Selection

The selection operation in the NSGA-II algorithm is performed using a special method called *crowded tournament selection*. In this method, two individuals i_1 and i_2 are randomly chosen from the population, and the winner is added to the mating pool. The selection is based on both feasibility and dominance criteria. The individual i_1 is wins over i_2 under the following conditions:

- i_1 is **feasible** and i_2 is **infeasible**.
- i_1 and i_2 are both feasible or unfeasible, but:
  - i_1 has a **lower rank** than i_2.
  - or i_1 and i_2 have the **same rank**, but i_1 has a **higher crowding distance**.

This approach ensures that feasible individuals are always preferred over infeasible ones, and diversity in the population is remained. The crowded tournament selection process used to build the mating pool is summarized in Algorithm 2.

Algorithm 2 Crowded Tournament Selection

Input: Population P, Mating pool size M
Output: Mating pool M_t
$M_t \leftarrow \emptyset$
while $|M_t| < M$ **do**
  Randomly select i_1, i_2 from P
  if `IsWinner`(i_1, i_2) **then**
    Add i_1 to M_t
  else
    Add i_2 to M_t
  end if
end while

C. Variation

By applying crossover and mutation procedures to the individuals in the mating pool, the genetic algorithm's variation stage produces new offspring. This implementation uses a combination of *Simulated Binary Crossover (SBX)* and a customized *random mutation* operator, both were modified to take biological factors and resource limitations into consideration.

First, the mating pool is shuffled, and individuals are paired. Each pair undergoes variation operation including a possibility of crossover and mutation. If crossover is applied (based on the crossover rate), offspring are generated using SBX applied to the `genetic_resources` and `ph_tolerance` parameters. Following crossover, biological parameters (e.g., DFI, motility, morphology, and velocity) are recalculated based on the updated resource allocation. Additionally, HLA profiles are recombined while taking into account the limitations of genetic resources.

Then by applying bounded random changes to resource allocation, pH tolerance, biological parameters, and HLA profiles (based on the crossover rate) to the offspring, the mutation is performed. Like the crossover operator, the mutation operator ensures that changes respect bounds. Small variations are also applied to biological parameters using a noise factor to mimic realistic variation.

Finally, the new offspring Q_t is generated. The process is summarized in Algorithm 3.

D. Survivor

The survivor stage in the NSGA-II algorithm determines which individuals are passed to the next generation. It begins

by merging the current population P with the offspring population Q_t. This combined population (R_t) is re-evaluated using non-dominated sorting and crowding distance assignment, as described in Section 1.

Next, the *rank filtering* step is performed to select the top N individuals based on their dominance rank and crowding distance. Individuals from the lowest-rank fronts are selected first. If the size of the front is bigger than the remaining empty slots for the next population, only the individuals with the highest crowding distances from that front are selected to preserve diversity. The process is summarized in Algorithm 4.

Algorithm 3 Variation Operator

Input: Mating pool M, crossover rate $\mathcal{C}$, mutation rate $\mathcal{M}$
Output: Offspring population Q_t
$Q_t \leftarrow \emptyset$
ParentPairs $\leftarrow$ `CreateParentPairs`(M)
for each (p_1, p_2) in ParentPairs **do**
  if random() $< \mathcal{C}$ **then**
    $(c_1, c_2) \leftarrow$ `SBXCrossover`(p_1, p_2)
  else
    $(c_1, c_2) \leftarrow$ `CopyParents`(p_1, p_2)
  end if
  `Mutate`(c_1, $\mathcal{M}$), `Mutate`(c_2, $\mathcal{M}$)
  $Q_t \leftarrow Q_t \cup \{c_1, c_2\}$
end for

Algorithm 4 Survivor Selection

Input: Parent population P, Offspring population Q_t, Population size N
Output: Next generation population P_t
$R_t \leftarrow P \cup Q_t$
$F \leftarrow$ `Evaluate`(R_t) // see Algorithm 1
$P_t \leftarrow \emptyset$
$n \leftarrow 0$
for each front $F_k \in F$ **do**
  if $n + |F_k| \leq N$ **then**
    Add all individuals in F_k to P_t
    $n \leftarrow n + |F_k|$
  else
    Sort F_k by descending crowding distance
    Add top $(N - n)$ individuals from F_k to P_t
    break
  end if
end for

IV. Experiments

This section presents a series of experiments conducted to evaluate the performance and behavior of the proposed multi-objective genetic algorithm under various parameter settings. The goal is to observe how different configurations affect the convergence, diversity, and feasibility of the generated solutions.

A. Parameters

To test the robustness and sensitivity of the algorithm, several key parameters are varied systematically:

- **Number of generations:** {100, 200, 500}
- **Population size:** {25, 50, 100}
- **Crossover rate ($\mathcal{C}$):** {0.6, 0.9}
- **Mutation rate ($\mathcal{M}$):** {0.05, 0.1}
- **SBX distribution index (η):** {5, 15}

B. Outputs

This section shows the output of the algorithm for reference. It is run with a random parameter combination. The initial population is created as Figure 1. Then after a certain number of generations, the final population becomes like Figure 2. The results show that the algortihm successfuly converged to pareto-front

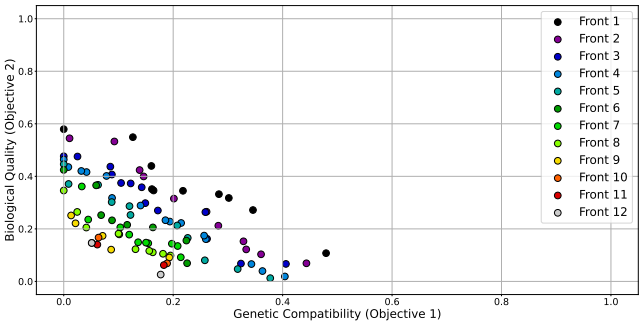


Fig. 1: Initial population with different pareto-front levels

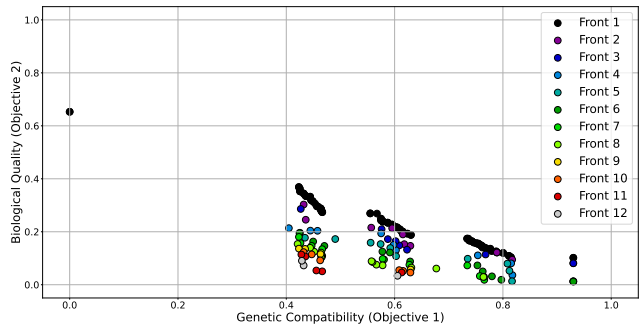
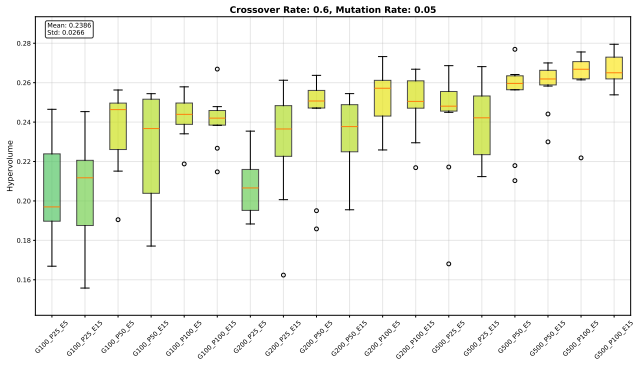


Fig. 2: Final population with different pareto-front levels

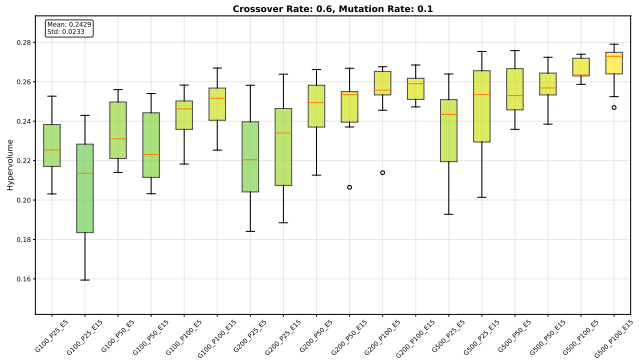
C. Hypervolume Indicator

Each parameter combination is run ten times to reduce the impact of randomnes. The *hypervolume indicator* is used to evaluate the performance of each configuration, as it reflects both the convergence and diversity of the obtained Pareto front.

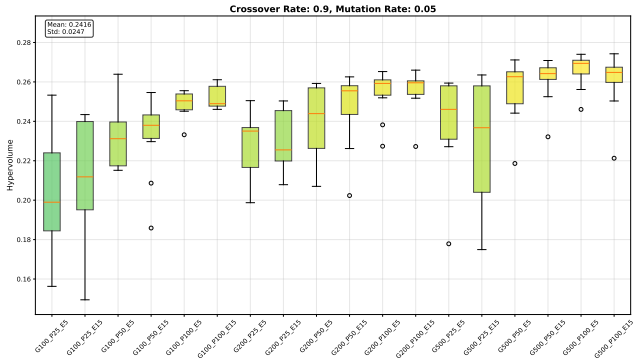
The results across all runs and parameter sets are visualized in Figure 3, showing how different configurations influence optimization performance.



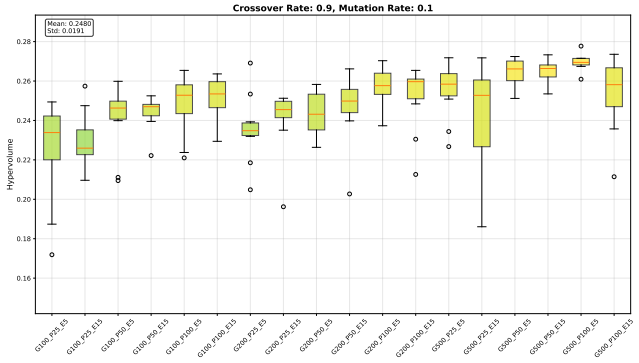
(a) Part 1/4 for Hypervolume plots



(b) Part 2/4 for Hypervolume plots



(c) Part 3/4 for Hypervolume plots

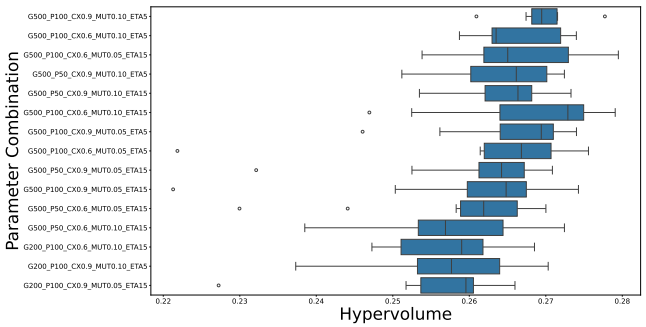


(d) Part 4/4 for Hypervolume plots

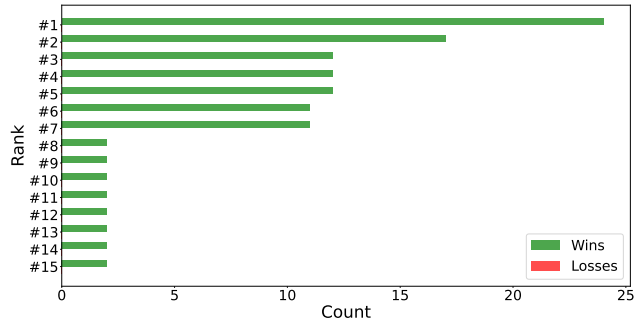
Fig. 3: Comparison of hypervolume scores across different parameter combinations. Each subfigure represents a specific **crossover-mutation** configuration along with different combinations for **number of generation**, **population size**, and **eta**.

D. Statistical Analysis

In order to identify the optimal parameter settings, statistical analysis is performed. Initially, one-way ANOVA test is used to determine whether variations in hypervolume scores among parameter combinations were statistically significant. To determine which combinations performed better than others, pairwise comparisons were conducted using independent t-tests with Bonferroni correction after such differences were discovered. After that, each configuration was ranked according to its mean hypervolume score and statistical win-loss balance. The results of the statistical analysis including the best fifteen parameter combinations are presented in Figure 4



(a) Hypervolume indicators for top fifteen parameter combinations



(b) Win/loss counts for each parameter combination against other parameters

Fig. 4: Top fifteen parameter combinations in terms of hypervolume indicator (a) and their winn/loss counts against all the combinations (b). The combinations are sorted from best to worth, ranks in (b) indicating their performance ranks.

V. Discussion

This work presents a biologically inspired multi-objective optimization framework for simulating egg-sperm compatibility using NSGA-II. By modeling key biological features such as HLA profile, DNA fragmentation index (DFI), motility, morphology, velocity, and pH tolerance; the trade-off between biological quality and genetic compatibility is effectively captured by the algorithm. The limited resource-based decision variables generated realistic constraints that lead to diverse and meaningful Pareto fronts.

Addition of dynamic penalty function ensures that infeasible individuals are vanished as the generations increase.

Additionally, the use of crowded tournament selection and rank-based survivor filtering helps maintaining the population diversity.

Experimental results and statistical analysis show how differently the algorithm behaves with different parameter combinations. Larger populations and higher crossover rates generally led to better hypervolume scores, as well as high generation counts.

VI. Conclusion

This study presents a NSGA-II frame-worked multi-objective optimization approach for egg-sperm compatibility simulation. Simplified yet representative abstraction of the problem is created for this project and by experimentally exploring the different parameter combinations, successful results are obtained. One-way ANOVA and Bonferroni tests are conducted to statistically select the top-performing parameter combinations. This study creates a fundamental structure for multiobjective evolutionary algorithms on different compatibility scenarios such as such as in vitro fertilization treatment (IVF).

References

[1] G. Myles and J. Primakoff, "The Fertilization Enigma: How Sperm and Egg Fuse," *Annual Review of Cell and Developmental Biology*, vol. 36, pp. 403–430, 2020. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-cellbio-120219-021751

[2] A. Jiménez-Movilla and M. Avilés, "Dynamics and Mechanisms of Sperm-Egg Interaction," *Frontiers in Cell and Developmental Biology*, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcell.2021.657037/full

[3] J. L. Fitzpatrick and M. J. Wolfner, "Genetic Incompatibility of Reproductive Partners: An Evolutionary Perspective," *PNAS*, vol. 118, no. 50, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8600657/

[4] J. L. Fitzpatrick et al., "Chemical signals from eggs facilitate cryptic female choice in humans," *Proceedings of the Royal Society B*, vol. 287, no. 1925, 2020. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rspb.2020.0805

[5] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[6] P. Neves, C. Nunes, J. M. Teixeira, J. F. Oliveira, and A. Viana, "Human Leukocyte Antigen Alleles Compatibility and Immunophenotypic Profile Associations in Infertile Couples," *Cureus*, vol. 15, no. 2, pp. e34627, Feb. 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10035384/

[7] J. Zhou, X. Chen, W. Li, Y. Song, and H. Wang, "The Semen pH Affects Sperm Motility and Capacitation," *PLOS ONE*, vol. 10, no. 7, p. e0132974, Jul. 2015. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132974

[8] X. Liu et al., "Sperm DNA fragmentation is positively correlated with miscarriage rate and negatively correlated with birth weight in IVF/ICSI cycles," *Scientific Reports*, vol. 13, no. 1, p. 18578, Oct. 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-45091-6

[9] R. F. Aziz et al., "Assessment of semen parameters and sperm DNA integrity in infertile men," *Middle East Fertility Society Journal*, vol. 27, no. 1, 2022. [Online]. Available: https://mefj.springeropen.com/articles/10.1186/s43043-022-00094-7

[10] A. Mostafa and H. ElBaz, "Sperm morphology and its impact on male fertility," *Frontiers in Reproductive Health*, vol. 4, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frph.2022.945351/full

[11] J. L. Fitzpatrick et al., "Sperm swimming velocity predicts competitive fertilization success in the guppy," *PLOS ONE*, vol. 5, no. 8, p. e12146, Aug. 2010. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012146

[12] "Can Sperm and Egg Be Genetically Incompatible?," Remembryo, [Online]. Available: https://www.remembryo.com/can-sperm-and-egg-be-genetically-incompatible/

[13] A. La Marca et al., "Ovarian reserve markers and IVF outcome: a meta-analysis," *Human Reproduction Update*, vol. 27, no. 1, pp. 48–63, 2021.