Received January 3 2024, accepted June 9 2025, date of publication September 2 2025.

Original Research Article

Characterization of Odor Profiles Through the Simplified Binary Matching Algorithm for Disease Diagnostics

Joanie B. Houinsou*, Roland C. Houessouvo, Kokou M. Assogba and Daton Medenou

Ecole Doctorale des Sciences de l'Ingénieur (ED-SDI)/Laboratoire LETIA/EPAC, Université d'Abomey Calavi (UAC), Abomey Calavi, 01 BP 2009 Cotonou, Bénin.

* Corresponding Author Email: jobhouinsou04@gmail.com

ABSTRACT

Background and Objective: This study investigates the characterization of body odor signatures for early disease detection, aiming to demonstrate the feasibility of using simulated olfactory profiles within a computational diagnostic framework. The motivation arises from the growing interest in non-invasive diagnostic alternatives based on volatile organic compounds (VOCs) emitted by the human body. **Materials and methods:** A simulation-based approach was implemented using validated VOC datasets to construct binary odor profiles. These profiles were encoded as binary vectors, with each bit indicating the presence or absence of a specific compound. A simplified binary matching algorithm, excluding mutation and crossover operations, was employed to simulate pattern matching. The Hamming distance was used as the fitness function to quantify the similarity between profiles. **Results and Discussion:** The results indicate that the simplified binary matching algorithm reliably identified pathological odor profiles, producing high similarity scores with reference signatures. Despite the absence of conventional genetic operators, the method consistently converged to optimal or near-optimal matches. These findings emphasize the potential of binary odor encoding for distinguishing between healthy and pathological states, underscoring the robustness of the simplified computational framework. **Conclusion:** This work presents a novel and interpretable computational model for olfactory-based disease detection using simulated binary VOC patterns. It supports the development of low-cost, non-invasive diagnostic tools in medical contexts. Future research should explore extending the method by incorporating continuous VOC encoding, integrating evolutionary operators, and validating the results with semi-experimental or clinical data.

Keywords—*Body odors, Diseases, Early detection, computational diagnostics, Simplified binary matching algorithm, Volatile organic compounds (VOCs).*

INTRODUCTION

The improvement of early disease detection methods is of crucial importance for public health.[1] Traditional approaches to medical diagnosis may be limited by cost, accessibility, and reliability. In this context, research is increasingly focusing on innovative methods based on artificial intelligence to enhance disease detection and prevention.[2] The genetic algorithm, an artificial intelligence technique inspired by the principles of natural evolution, holds promising potential in the field of early disease detection. By leveraging the adaptive and evolutionary capabilities of living organisms, this algorithm optimizes solutions for complex problems.[3]

In this context, this article examines the application of the genetic algorithm for early disease detection, with a specific focus on analyzing body odors composed of volatile organic compounds (VOCs). Recent research suggests that certain diseases can alter specific odor profiles of the human body,[4] providing an opportunity to use olfactory information as an early indicator of health issues.

The objective of this article is to present a methodology based on a simplified pattern-matching algorithm, inspired by the principles of genetic algorithms, for analyzing body odors and detecting diseases at an early stage. Unlike conventional genetic algorithms that incorporate selection, crossover, and mutation operations, the method implemented here deliberately omits these evolutionary components. Instead, it evaluates binary-encoded odor profiles using Hamming distance to identify the closest match to a target profile. This simplification aims to enhance interpretability, reproducibility, and computational efficiency within a purely simulation-based framework. By integrating expertise in genetics, artificial intelligence, and medicine, this approach could contribute to revolutionizing medical diagnostic methods by enabling faster, more accurate, and less invasive disease detection.

LITERATURE REVIEW

Body Odors

Volatile organic compounds (VOCs) are chemical substances released by the human body, playing a significant role in body odor.[5,6] In ancient societies, body odors were more prevalent and accepted, regarded as part of individual and social identity. Over time, attitudes toward body odor have evolved alongside scientific advances and social norms.[7] In medieval Europe, body odor was associated with notions of sin and decadence due to religious beliefs, and perfumes were commonly used to mask undesirable odors. In the modern era, hygiene and cleanliness became priorities, leading to the development of personal care products to control body odor. However, these products may alter natural odors by adding fragrances.[8] Recently, certain movements have advocated for the acceptance of natural body odor, and challenged social norms that aim to eliminate it. Body odors vary among individuals due to various factors, and perceptions of body odor differ across cultures.[9]

Early Detection of Diseases

The early detection of diseases plays a crucial role in preserving health and well-being. It enables prompt intervention by identifying early signs and symptoms, leading to more favorable outcomes in terms of treatment, management, and even cure.[10] Numerous benefits are associated with the early detection of diseases.[11] Firstly, it allows for rapid medical intervention, helping prevent disease progression and reduce potential complications.[12] Secondly, it increases the likelihood of treatment success, as interventions are often more effective when administered at an early stage of the disease. Additionally, it helps reduce long-term healthcare costs, as early treatments are typically less invasive and less expensive than those required at an advanced stage of the disease.[13]

To detect diseases early, various methods are employed. Regular screenings and health examinations are essential for identifying early signs of common diseases such as breast, cervical, and colon cancers. Technological advancements have led to the development of sophisticated blood tests and medical imaging techniques, which can aid in detecting diseases at an early stage, even in the absence of apparent symptoms.[14] Furthermore, genetics and personalized medicine have opened new possibilities by identifying genetic markers associated with certain conditions.[15]

Early disease detection helps limit potential complications, improve patients' quality of life, and reduce the burden on healthcare systems.[12,16]

Genetic Algorithm

Genetic Algorithms (GAs) were first described by John Holland in the 1960s and later developed by him, his students, and colleagues at the University of Michigan during the 1960s and 1970s. Holland's objective was to understand the phenomenon of adaptation as it occurs in nature and to develop methods for incorporating the mechanisms of natural adaptation into computer systems.[17]

The genetic algorithm (FIGURE 1) is a computational approach inspired by the process of biological evolution. It is a search and optimization method based on the principles of natural selection and genetics. Genetic algorithms are widely applied to solve complex problems in various fields, including engineering, optimization, artificial intelligence, and bioinformatics.[18,19]

The genetic algorithm operates by simulating an artificial evolution process, in which an initial population of individuals (often represented by bit strings) is randomly generated. Each individual in the population is evaluated based on its performance relative to a specific goal defined by an evaluation function.[20]

The crucial step in the genetic algorithm is selection. The fittest individuals, i.e., those with the best performance, are chosen to reproduce and produce offspring. This selection is typically based on a method called "fitness-proportional selection", in which the probability of selection is proportional to the fitness value of each individual.[17,21,22]

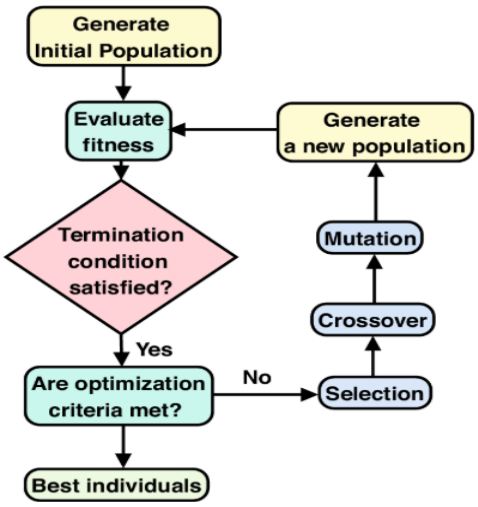

FIGURE 1. General workflow of a canonical genetic algorithm.[23]

Once selection is complete, genetic operations are applied to the offspring. These operations include recombination (crossover) and mutation. Recombination involves combining the genetic information of two selected individuals to create new individuals, while mutation introduces random changes in individuals to explore new potential solutions.[17,21,22] This process of selection, recombination, and mutation is repeated over several generations, allowing the population to gradually converge toward increasingly optimal solutions. The genetic algorithm may also incorporate techniques such as elitism, which involves retaining the best individuals from one generation to the next to ensure faster convergence.[17,21,22]

MATERIALS AND METHODS

Materials

We utilized the Human Metabolome Database (HMDB) to obtain detailed information on small-molecule metabolites present in the human body. The objective was to apply this information for biomarker discovery applications.[24,25]

Additionally, we relied on the work of reference[26], which cataloged over 1,800 volatile organic compounds emitted by the body of a healthy individual. We also consulted the Cancer Odor Database (COD), an online resource documenting known volatile organic metabolites of cancer (VOMC), commonly referred to as "cancer odors".[27]

Finally, Broza et al. [28] highlights the increasing importance of developing new diagnostic and detection technologies to address growing clinical challenges. It emphasizes a new diagnostic frontier based on detecting disease-associated volatile organic compounds (VOCs) using sensors that employ nanomaterials.

Population Dataset Construction

To construct the initial population used in our simulations, we compiled a comprehensive dataset of 2,571 volatile organic compounds (VOCs) from authoritative sources, including the Human Metabolome Database (HMDB), the Cancer Odor Database (COD), and peer-reviewed literature, such as the catalog published by.[26] Each VOC was annotated with a unique CAS number and labeled according to its known association with physiological or pathological states, including various cancer types and healthy conditions.

This dataset served as the basis for generating 25 distinct binary vectors, each representing a specific simulated odor signature associated with a defined condition. The encoding process involved mapping the presence (1) or absence (0) of each of the 2,571 VOCs for every condition, resulting in uniform-length binary chromosomes. These chromosomes were stored and processed as the initial population from which the algorithm searched for the best match to a given target.

The structured nature and dimensional richness of this population enabled meaningful comparison and pattern recognition through Hamming distance evaluation. Importantly, the dataset was constructed to balance diversity (in terms of represented conditions) and consistency (in binary structure), ensuring that the algorithm operated within a representative yet tractable search space.

Methods

The method employed for disease detection is the simplified binary matching algorithm, which encompasses five phases.

Phase 1: Individual Representation

Each individual, denoted by Equation 1, is symbolically characterized by a chromosome—a structured sequence of fixed-length binary digits that corresponds to the quantity of volatile organic compounds (VOCs) defining the olfactory profile. The chromosome is mathematically expressed as:

$$Chromosome_i = [v_{i1}, v_{i2}, \ldots, v_{in}] \quad (1)$$

Within the confines of this representation (Equation 1), each element v_{ij} (Equation 2) is discretized into a binary bit, serving as an indicator of the presence or absence of a specific VOC. This binary encoding is represented by the formula:

$$v_{ij} = \begin{cases} 0, If\ the\ VOC\ is\ not\ present \\ 1, If\ the\ VOC\ is\ present \end{cases} \quad (2)$$

where $v_{i1}, v_{i2}, \ldots, v_{in}$, represent the elements of the chromosome for sample i, v_{ij} denotes the presence (1) or absence (0) of the j_th VOC, n is the total number of VOCs considered, and $i=1,2, \ldots, N$ indexes the sample.

Consequently, the collective exposition of equations (Equation 1) and (Equation 2) coherently explicates the chromosome's nature as a combination of binary units, delineating the presence or absence of VOCs within the context of individual representation. v_{ij} is the binary variable indicating the state of the j_th VOC in the i_th chromosome.

Phase 2: Evaluation Function

An evaluation function assigns a value, or fitness score, to each chromosome based on its ability to solve the given problem. In this study, the Hamming function serves as the objective function (Equation 3), calculating the distance between the desired solution and the candidate solution within the population.

$$f\left(Chromosome_i, Chromosome_s\right) = -\sum_{j=0}^{N-1} Abs(Chromosome_{ij} - Chromosome_{sj}) \quad (3)$$

where $Chromosome_s$ denotes the binary vector representing the i_th individual (candidate solution) in the population, $Chromosome_s$ the target chromosome corresponding to the reference (disease) profile, and f ($Chromosome_s$, $Chromosome_s$) the fitness function measuring their similarity. The term Abs($Chromosome_{ij}$, $Chromosome_{sj}$)defines the absolute difference between the candidate and the target at the position j , while the summation counts $\sum_{j=0}^{N-1} Abs(Chromosome_{ij} - Chromosome_{sj})$ the total number of mismatches between the two binary vectors.

In our implementation, the fitness score is defined as the negative Hamming distance between the target profile and each candidate in the population. This transformation (multiplication by −1) enables the interpretation of higher scores—values closer to zero—as better matches, while preserving the relative ranking of similarity. A score of 0 represents a perfect match, whereas increasingly negative scores indicate greater dissimilarity.

Phase 3: Population Initialization

We initialized the population with binary data imported from a specially prepared Excel file, following the methodology described in previous studies.[24–26,28] These data define the search space for disease identification, representing the presence or absence of volatile organic compounds (VOCs) associated with specific conditions

Phase 4: Main Loop of the Algorithm

The main loop of the algorithm concluded when the predefined termination criterion was met, which in our case was a fixed number of iterations equal to the population size.

Phase 5: Results Analysis

This phase involves examining the individuals to identify those that correspond to the best solution found and extracting relevant information from them to address our problem.

FIGURE 2 illustrates the workflow of a simplified pattern-matching algorithm that identifies the binary individual within a given population that best matches a predefined target profile.

The procedure begins by initializing variables to store the best-known match, then iteratively evaluates the Hamming distance between each candidate chromosome and the target. Whenever a closer match is identified, the best candidate is updated. The algorithm concludes by returning the individual with the smallest distance to the target. This approach is deterministic, easily interpretable, and does not employ stochastic genetic operators such as crossover or mutation.

This flowchart (FIGURE 2) illustrates the sequence of steps in the proposed deterministic algorithm:

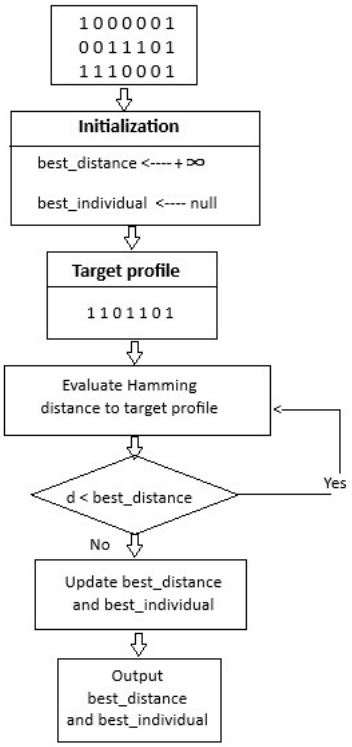

FIGURE 2. Workflow of the simplified binary matching algorithm.

1. Initialization of the binary-encoded population based on VOC presence/absence;

2. Comparison of each individual with a target profile using the Hamming distance;

3. Selection of the profile with the minimum distance as the optimal match.

The algorithm bypasses traditional genetic operations—such as selection, mutation, and crossover—and relies exclusively on distance-based evaluation.

FIGURE 3 presents the pseudocode for a simplified computational procedure designed to identify the individual within a binary population that most closely matches a given target profile. The method iteratively computes the Hamming distance between each chromosome and the target, updating the best match whenever a smaller distance is encountered. This approach enables efficient nearest-neighbor selection in discrete binary spaces while eliminating the need for evolutionary operators such as crossover or mutation.

RESULTS AND DISCUSSION

Results

The simplified binary matching algorithm, applied to early disease detection through body odor analysis, produced the following results:

Presentation of Data in Binary Form

We obtained data encoded in binary form, as illustrated in Figure 4, representing our search space.

```
Input: Population P = {Chromosome_1, ..., Chromosome_n}, Target_Profile
Best_Distance ← +∞
Best_Individual ← None

for each Chromosome_i in P:
    d ← Hamming_Distance(Chromosome_i, Target_Profile)
    if d < Best_Distance:
        Best_Distance ← d
        Best_Individual ← Chromosome_i

Output: Best_Individual, Best_Distance
```

FIGURE 3. Pseudocode for a simplified matching algorithm based on hamming distance.

	CAS-number	Compound name	Codes	Feaces	Urine	Breath	Skin	Milk	Blood	Saliva	...	Pancreatic_Cancer	Prostate_Cancer	Skin_Cancer	Synovial_Cancer	T
0	75-07-0	acetaldehyde	v1	1	1	1	1	1	1	1	...	0	0	0	0	
1	60-35-5	acetamide	v2	1	0	1	0	0	0	0	...	0	0	0	0	
2	64-19-7	acetic acid	v3	1	1	1	1	1	0	1	...	0	0	0	0	
3	140-11-4	acetic acid, benzyl ester	v4	0	0	0	1	0	0	1	...	0	0	0	0	
4	123-86-4	acetic acid, butyl ester	v5	1	0	1	0	0	0	0	...	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2566	64-17-5	ethanol	v2567	0	0	0	0	0	0	0	...	0	0	0	0	
2567	598-58-3	methyl nitrate	v2568	0	0	0	0	0	0	0	...	0	0	0	0	
2568	598-58-3	Methyl nitrate	v2569	0	0	0	0	0	0	0	...	0	0	0	0	
2569	1330-20-7	xylene	v2570	0	0	0	0	0	0	0	...	0	0	0	0	
2570	100-41-4	ethylbenzene	v2571	0	0	0	0	0	0	0	...	0	0	0	0	

2571 rows × 30 columns

FIGURE 4. Binary encoding of VOCs related to body odor.

Best Individual and Fitness Score

The code outputs the best individual identified within the population of potential solutions—namely, the body odor sequence achieving the highest fitness score. This score, derived from the Hamming distance between the individual and the target sequence, facilitates the identification of the most effective profiles for disease detection based on body odor. Execution traces of the algorithm are presented in Figures 5–12.

Processing by our simplified binary matching algorithm on a randomly generated chromosome classified it as healthy, with a fitness score of −1,257 (FIGURE 5), indicating a high similarity to the healthy reference profile. The corresponding fitness scores for this case are shown in FIGURE 6.

Processing by the same algorithm on another randomly generated chromosome classified it as diseased, with a fitness score of −1,241 (FIGURE 7), indicating slightly lower similarity to the healthy reference profile. The corresponding fitness scores are presented in Figure 8. Within our search space, this chromosome is associated with breast cancer.

A diseased chromosome from the search space was processed using our simplified binary matching algorithm, which identified it with a fitness score of 0 (FIGURE 9).

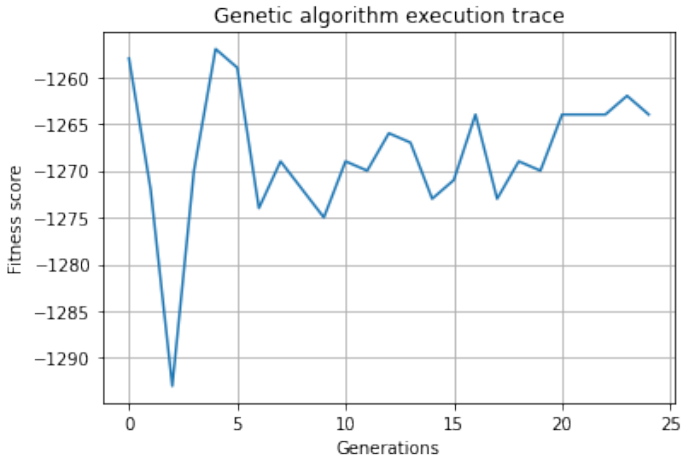

FIGURE 6. Fitness scores generated by a random healthy chromosome.

In our search space, this chromosome corresponds to an individual with neck cancer (FIGURE 10).

Regardless of the target chromosome, the simplified binary matching algorithm converges toward an optimal solution with an associated score. A score of 0 (FIGURE 1) indicates that the target chromosome is present in the solution space; otherwise, the algorithm identifies the chromosome in the space that is closer to the target than any other (FIGURE 2). The chromosome in the

```
target_sequence [0 1 0 ... 1 0 1]
Best individual : [1 0 1 ... 0 0 0]
Fitness score : -1257
Index found : 4
fitness_scores [-1258, -1272, -1293, -1270, -1257, -1259, -1274,
-1269, -1272, -1275, -1269, -1270, -1266, -1267, -1273, -1271,
-1264, -1273, -1269, -1270, -1264, -1264, -1264, -1262, -1264]
The closest line :
CAS-number                  Milk
75-07-0                        1
60-35-5                        0
64-19-7                        1
140-11-4                       0
...
64-17-5.4                      0
598-58-3                       0
598-58-3.1                     0
1330-20-7                      0
100-41-4.2                     0
Name: 6, Length: 2572, dtype: object
```

FIGURE 5. Algorithm trace for a random healthy chromosome.

```
target_sequence [1 0 1 ... 1 0 1]
Best individual : [0 0 0 ... 0 0 0]
Fitness score : -1241
Index found : 7
fitness_scores [-1296, -1284, -1301, -1322, -1291, -1277, -1288,
 -1241, -1284, -1265, 1279, -1274, -1274, -1277, -1269, -1269,
-1274, -1273, -1273, -1282, -1272, -1272, -1272, -1268, -1272]
The closest line :
CAS-number              Breast_Cancer
75-07-0                             0
60-35-5                             0
64-19-7                             0
140-11-4                            0
                                  ...
64-17-5.4                           0
598-58-3                            0
598-58-3.1                          0
1330-20-7                           0
100-41-4.2                          0
Name: 9, Length: 2572, dtype: object
```

FIGURE 7. Algorithm trace for a random diseased chromosome.

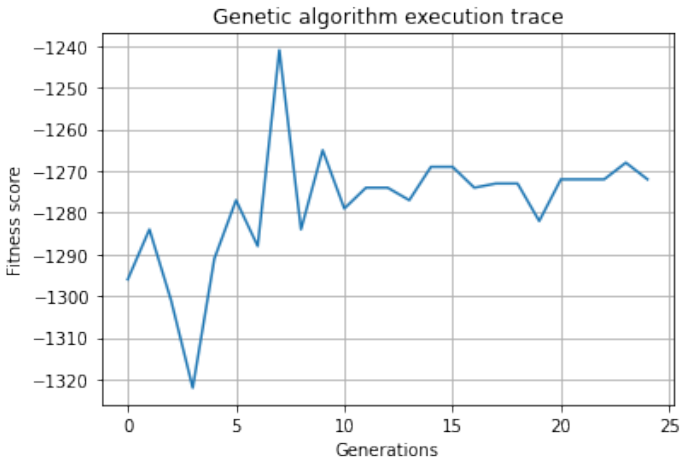

FIGURE 8. Fitness scores generated by a random diseased chromosome.

```
target_sequence [0 0 0 ... 0 0 0]
Best individual : [0 0 0 ... 0 0 0]
Fitness score : 0
Index found : 10
fitness_scores [-393, -291, -886, -543, -268, -168,
 -371, -94, -73, -62, 0, -35, -31, -42, -286, -28,
-17, -18, -20, -69, -23, -31, -45, -23, -15]
The closest line :
CAS-number     Head_Neck_Cancer
75-07-0                       0
60-35-5                       0
64-19-7                       0
140-11-4                      0
                    ...
64-17-5.4                     0
598-58-3                      0
598-58-3.1                    0
1330-20-7                     0
100-41-4.2                    0
Name: 12, Length: 2572, dtype: object
```

FIGURE 9. Algorithm trace for a selected diseased chromosome in the search space.

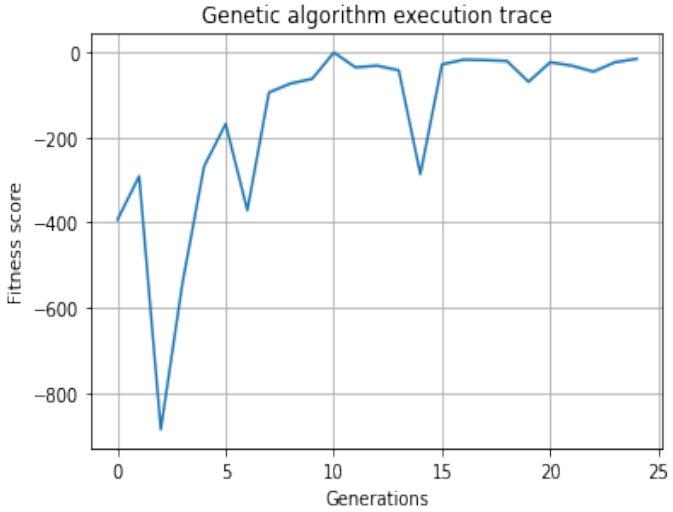

FIGURE 10. Fitness scores generated by a selected diseased chromosome with neck cancer in the search space.

space that is closer to the target than any other (FIGURE 2).Regardless of the target chromosome, the simplified binary matching algorithm converges toward an optimal solution with an associated score. A score of 0 (FIGURE 1) indicates that the target chromosome is present in the solution space; otherwise, the algorithm identifies the chromosome in the space that is closer to the target than any other (FIGURE 2).

Index of the Nearest Data Point

The code identifies the index of the nearest data point in the fitness score sequence. This index is then used to associate the corresponding data with additional information for results analysis. In our case, it links to detailed information about the volatile organic compounds associated with either a diseased or a healthy individual, as determined by their unique CAS identification number.

Fitness Scores

The code outputs the list of fitness scores for each individual in the population at every generation. This enables visualization of score evolution over time and facilitates tracking of the algorithm's progress in identifying the best individual. As the search space becomes more enriched, these scores are expected to improve.

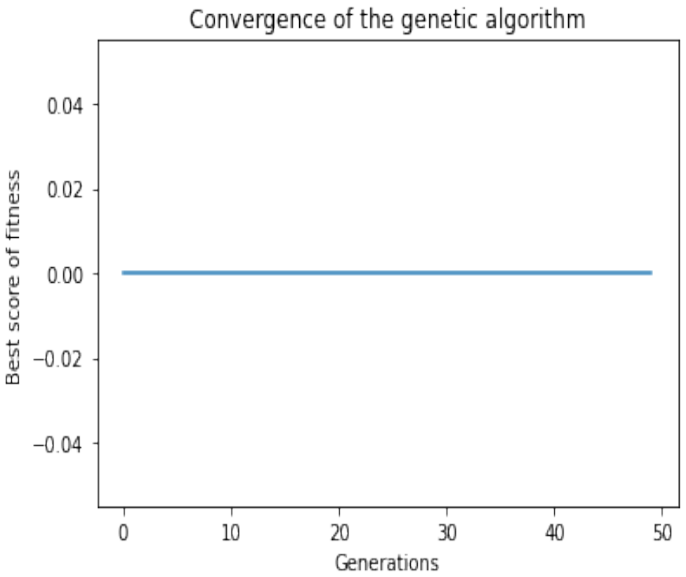

FIGURE 11. Convergence to zero of the algorithm for a target chromosome present in the search space.

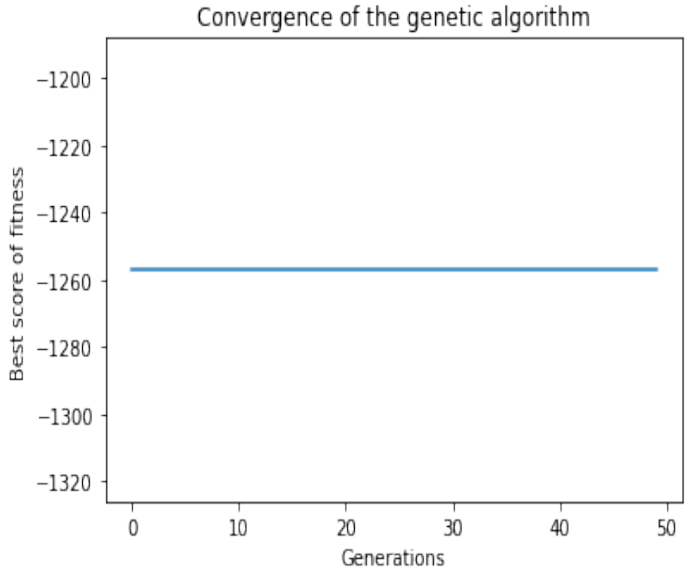

FIGURE 12. Convergence to a finite score of the algorithm for a target chromosome absent in the search space.

Nearest Line

The code retrieves information associated with the nearest data point from the Excel file used to generate the initial population. This allows for examining the specific details of that data and analyzing them in relation to the results of the simplified binary matching algorithm.

Algorithm Convergence

Regardless of the target chromosome, the algorithm converges towards an optimal solution.

DISCUSSION

By deliberately omitting evolutionary components such as selection, crossover, mutation, and replacement, the algorithmic procedure in this study deviates from conventional genetic algorithms, adopting instead a deterministic, pattern-matching framework. The resulting model functions solely through the initialization and evaluation of a predefined population of binary VOC profiles. Each individual in the population represents a potential solution encoded as a fixed-length binary vector, with evaluation performed using the Hamming distance as the fitness metric.

This simplified configuration enhances interpretability and reproducibility by avoiding the stochastic variability and convergence dynamics inherent in evolutionary systems. Although this design sacrifices the exploratory capabilities of classical genetic algorithms, it is well-suited for simulation scenarios in which the search space is predefined and fully enumerable.

To ensure that the dataset retained discriminatory power despite the absence of evolutionary mechanisms, we conducted a distributional analysis using the Hamming distance metric. This analysis assessed profile diversity and spatial separability within the binary encoding space. The results confirmed that the initial population preserved sufficient structural variability to support meaningful pattern recognition.

CONCLUSION

In conclusion, the simplified binary matching algorithm demonstrates significant potential for early disease detection based on body odors within medical diagnostics. Analysis of volatile organic compounds present in body odor offers valuable insights into an individual's health status. The study's results indicate a strong correlation between body odor profiles, volatile organic compounds, and disease presence, thereby opening new avenues for non-invasive and cost-effective diagnostic methods.

However, further studies and the establishment of standardized protocols are essential to validate this approach and ensure its clinical reliability. While the results

demonstrate the feasibility of body odor analysis for early detection of various diseases, additional research is necessary to improve the specificity and sensitivity of the method.

Integrating the simplified binary matching algorithm into body odor analysis offers the potential to optimize early disease detection, enabling faster and more effective medical intervention. Additionally, this non-invasive approach may enhance patient acceptance and participation.

Although inspired by the genetic algorithm paradigm, the implemented model diverges from traditional evolutionary computation by adopting a deterministic, non-stochastic structure. For clarity, the term "simplified binary matching algorithm" is used to reflect both its origins and methodological constraints.

Overall, the use of the simplified binary matching algorithm for early disease detection based on body odors presents promising new prospects in the medical field. This approach has the potential to improve treatment success rates by enabling early and accurate disease diagnosis.

AUTHOR CONTRIBUTIONS

Conceptualization, J.H. and K.A.; Methodology, J.H.; Software, J.H.; Hardware, J.H. and R.H.; Validation, J.H., R.H., and D.M.; Formal Analysis, J.H.; Investigation, J.H.; Resources, J.H.; Data Curation, J.H.; Writing–Original Draft Preparation, J.H.; Writing–Review & Editing, R.H. and D.M.; Visualization, J.H.; Supervision, K.A.; Project Administration, K.A.; Funding Acquisition, D.M.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

Not applicable.

CONFLICTS OF INTEREST

The authors declare they have no competing interest.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

FURTHER DISCLOSURE

Not applicable.

REFERENCES

1. Srivastava, S. and Gopal-Srivastava, R. Biomarkers in Cancer Screening: A Public Health Perspective. *J Nutr*. 2002;132(8 Suppl):2471S–2475S. https://doi.org/10.1093/jn/132.8.2471S.
2. Maselli, G., Bertamino, E., Capalbo, C., et al. Hierarchical convolutional models for automatic pneumonia diagnosis based on X-ray images: new strategies in public health. *Ann Ig Med Prev E COMUNITÀ*. 2021;(6):644–655. https://doi.org/10.7416/ai.2021.2467.
3. Gen, M. and Cheng, R.W. *Genetic Algorithms and Engineering Optimization*. John Wiley & Sons: Hoboken, USA; 1999; 520 p.
4. Shirasu, M. and Touhara, K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J Biochem (Tokyo)*. 2011;150(3):257–266. https://doi.org/10.1093/jb/mvr090.
5. Smallegange, R.C., Verhulst, N.O., Takken, W. Sweaty skin: an invitation to bite? *Trends Parasitol*. 2011;27(4):143–148. https://doi.org/10.1016/j.pt.2010.12.009.
6. Aksenov, A.A., Gojova, A., Zhao, W., et al. Characterization of Volatile Organic Compounds in Human Leukocyte Antigen Heterologous Expression Systems: a Cell's "Chemical Odor Fingerprint". *ChemBioChem*. 2012;13(7):1053–1059. https://doi.org/10.1002/cbic.201200011.
7. Croijmans, I., Beetsma, D., Aarts, H., et al. The role of fragrance and self-esteem in the perception of body odors and impressions of others. *PLoS One*. 2021;16(11):e0258773. https://doi.org/10.1371/journal.pone.0258773.
8. Zakrzewska, M. Olfaction and prejudice: The role of body odor disgust sensitivity and disease avoidance in understanding social attitudes [thesis]. Stockholm University, Sweden, 2022.
9. Mutic, S., Moellers, E.M., Wiesmann, M., et al. Chemosensory communication of gender information: Masculinity bias in body odor perception and femininity bias introduced by chemosignals during social perception. *Front Psychol*. 2016;6:1980. https://doi.org/10.3389/fpsyg.2015.01980.
10. Danna, K., Griffin, R.W. Health and Well-Being in the Workplace: A Review and Synthesis of the Literature. *J Manag*. 1999;25(3):357–384. https://doi.org/10.1016/S0149-2063(99)00006-9.

11. Pereira, J., Porto-Figueira, P., Cavaco, C., et al. Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview. *Metabolites*. 2015;5(1):3–55. https://doi.org/10.3390/metabo5010003.
12. Etzioni, R., Urban, N., Ramsey, S., et al. The case for early detection. *Nat Rev Cancer*. 2003;3(4):243–252. https://doi.org/10.1038/nrc1041.
13. Akil, A.A., Yassin, E., Al-Maraghi, A., et al. Diagnosis and treatment of type 1 diabetes at the dawn of the personalized medicine era. *J Transl Med*. 2021;19(1):137. https://doi.org/10.1186/s12967-021-02778-6.
14. Haleem, A., Javaid, M., Singh, R.P., et al. Biosensors applications in the medical field: A brief review. *Sens Int*. 2021;2:100100. https://doi.org/10.1016/j.sintl.2021.100100.
15. Lee, Y. and Wong, D.T. Saliva: an emerging biofluid for early detection of diseases. *Am J Dent*. 2009;22(4):241–248.
16. Kang, S., Peng, W., Zhu, Y., et al. Recent progress in understanding 2019 novel coronavirus (SARS-CoV-2) associated with human respiratory disease: detection, mechanisms, and treatment. *Int J Antimicrob Agents*. 2020;55(5):105950. https://doi.org/10.1016/j.ijantimicag.2020.105950.
17. Mitchell, M. *An Introduction to Genetic Algorithms*. MIT Press: London, England; 1996. 221p.
18. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company: Boston, USA; 1989. 412p.
19. Bayer, S.E. and Wang, L. A genetic algorithm programming environment: Splicer. In Proceedings of the 1991 Third International Conference on Tools for Artificial Intelligence. San Jose, USA, November 1991, IEEE: Piscataway, USA, 1991. p. 138–139.
20. McCall, J. Genetic algorithms for modelling and optimisation. *J Comput Appl Math*. 2005;184(1):205–222. https://doi.org/10.1016/j.cam.2004.07.034.
21. Haupt, R,L. and Haupt, S.E. *Practical Genetic Algorithms*. John Wiley & Sons: Hoboken, USA; 2004.
22. Reeves, C.R. Genetic Algorithms. In *Handbook of Metaheuristics*. Gendreau M, Potvin JY, editors. Springer: Boston, USA; 2010; pp. 109–139. (International Series in Operations Research & Management Science; vol. 146).
23. Sallah, A., Alaoui, E.A.A., Tekouabou, S.C.K., et al. Machine learning for detecting fake accounts and genetic algorithm-based feature selection. *Data Policy*. 2024;6:e15. https://doi.org/10.1017/dap.2023.46.
24. Wishart, D.S., Feunang, Y.D., Marcu, A., et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46(D1):D608–617. https://doi.org/10.1093/nar/gkx1089.
25. Wishart, D.S., Guo, A., Oler, E., et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res*. 2022;50(D1):D622–631. https://doi.org/10.1093/nar/gkab1062.
26. de Lacy Costello, B., Amann, A., Al-Kateb, H., et al. A review of the volatiles from the healthy human body. *J Breath Res*. 2014;8(1):014001. https://doi.org/10.1088/1752-7155/8/1/014001.
27. Janfaza, S., Banan Nojavani, M., Khorsand, B., et al. Cancer Odor Database (COD): a critical databank for cancer diagnosis research. *Database*. 2017;2017:bax055. https://doi.org/10.1093/database/bax055.
28. Broza, Y.Y. and Haick, H. Nanomaterial-Based Sensors for Detection of Disease by Volatile Organic Compounds. *Nanomed*. 2013;8(5):785–806. https://doi.org/10.2217/nnm.13.64.