

The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.



**MUTABLE COMPOSITE FIREFLY ALGORITHM FOR GENE  
SELECTION IN MICROARRAY BASED CANCER  
CLASSIFICATION**

**MOHAMED NISPER FATHIMA FAJILA**



**MASTER OF SCIENCE (INFORMATION TECHNOLOGY)  
UNIVERSITI UTARA MALAYSIA  
2022**



Awang Had Salleh  
Graduate School  
of Arts And Sciences

Universiti Utara Malaysia

**PERAKUAN KERJA TESIS / DISERTASI**  
(*Certification of thesis / dissertation*)

Kami, yang bertandatangan, memperakukan bahawa  
(*We, the undersigned, certify that*)

MOHAMED NISPER FATHIMA FAJILA

calon untuk Ijazah

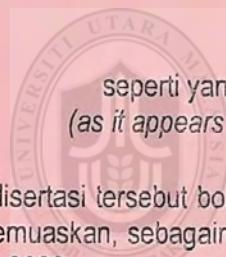
(*candidate for the degree of*)

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)

telah mengemukakan tesis / disertasi yang bertajuk:

(*has presented his/her thesis / dissertation of the following title*):

**"MUTABLE COMPOSITE FIREFLY ALGORITHM FOR GENE SELECTION IN MICROARRAY BASED CANCER CLASSIFICATION"**



seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.  
(*as it appears on the title page and front cover of the thesis / dissertation*).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **07 April 2022**.

*That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:*

**07 April 2022.**

Pengerusi Viva:  
(*Chairman for VIVA*)

Assoc. Prof. Dr. Siti Sakira Kamaruddin

Tandatangan  
(*Signature*)

Pemeriksa Luar:  
(*External Examiner*)

Assoc. Prof. Dr. Zalinda Othman

Tandatangan  
(*Signature*)

Pemeriksa Dalam:  
(*Internal Examiner*)

Assoc. Prof. Ts. Dr. Nor Hazlyna Harun

Tandatangan  
(*Signature*)

Nama Penyelia/Penyelia-penyalia: **Assoc. Prof. Dr. Yuhani Yusof**  
(*Name of Supervisor/Supervisors*)

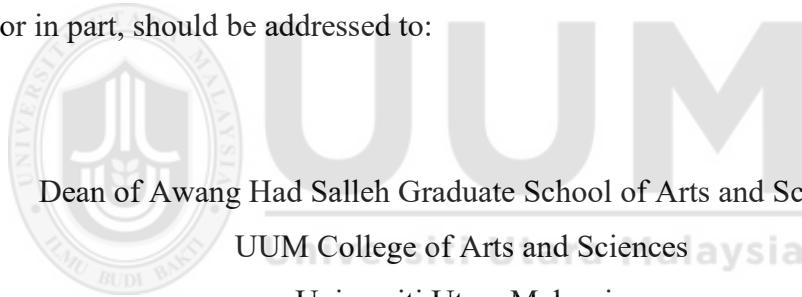
Tandatangan  
(*Signature*)

Tarikh:  
(*Date*) **07 April 2022**

## **Permission to Use**

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:



Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

## Abstrak

Pengelasan kanser adalah kritikal kerana keperluan usaha yang gigih dalam rawatan kanser dan peningkatan kadar kematian akibat kanser. Trend terkini dengan teknologi pemprosesan yang tinggi telah membawa kepada ciptaan baharu iaitu biopenanda yang berjaya menyumbang dalam penyelesaian isu berkaitan kanser. Pendekatan pengiraan untuk pemilihan gen dalam pengelasan kanser menggunakan data tatasusunan mikro telah diguna pakai dalam pelbagai masalah pengelasan kanser. Walau bagaimanapun, pendekatan hibrid sedia ada dengan algoritma pengoptimuman metaheuristik dalam pemilihan ciri (khususnya dalam pemilihan gen) tidak cukup umum untuk mengelasi kebanyakan data tatasusunan mikro kanser dengan cekap sambil mengekalkan set gen yang kecil. Ini membawa kepada permasalahan ketepatan pengelasan dan saiz subset gen. Oleh itu, kajian ini mencadangkan pengubahsuaian Algoritma Kunang-kunang (FA) bersama penapis Pemilihan Ciri (CFS) yang berasaskan korelasi dalam tugas pemilihan gen. Untuk mengatasi isu penumpuan perlahan yang dialami oleh FA, satu algoritma FA yang ditambah baik melalui penghasilan penyelesaian yang mempunyai saiz boleh ubah telah dicadangkan. Selanjutnya, satu strategi kemas kini kedudukan secara komposit telah direka bentuk untuk penyelesaian saiz boleh ubah. Strategi ini disasarkan untuk menangani masalah optimum tempatan dengan mengimbangi antara penerokaan dan eksplorasi dalam FA. Algoritma hibrid yang dicadangkan dikenali sebagai CFS-Algoritma Kunang-Kunang Boleh Ubah (CFS-MCFA) telah dinilai pada data tatasusunan mikro kanser untuk pemilihan biopenanda bersama-sama dengan penggunaan Mesin Sokongan Vektor (SVM) sebagai pengkelas. Penilaian dilakukan berdasarkan dua metrik; ketepatan pengelasan dan saiz set ciri. Keputusan menunjukkan bahawa algoritma CFS-MCFA-SVM menghasilkan prestasi yang baik dari segi ketepatan klasifikasi dan saiz subset gen berbanding kaedah penanda aras. Secara khususnya, ketepatan 100 peratus diperoleh pada keempat-empat set data dan dengan beberapa biopenanda (antara satu hingga empat). Hasil seperti ini menunjukkan algoritma yang dicadangkan ini adalah salah satu alternatif yang berdaya saing dalam pemilihan ciri, yang kemudiannya menyumbang dalam analisis data tatasusunan mikro.

**Kata kunci:** Pengelasan data, Algoritma Kunang-Kunang, Pemilihan gen, Data Tatasusunan Mikro.

## Abstract

Cancer classification is critical due to the strenuous effort required in cancer treatment and the rising cancer mortality rate. Recent trends with high throughput technologies have led to discoveries in terms of biomarkers that successfully contributed to cancer-related issues. A computational approach for gene selection based on microarray data analysis has been applied in many cancer classification problems. However, the existing hybrid approaches with metaheuristic optimization algorithms in feature selection (specifically in gene selection) are not generalized enough to efficiently classify most cancer microarray data while maintaining a small set of genes. This leads to the classification accuracy and genes subset size problem. Hence, this study proposed to modify the Firefly Algorithm (FA) along with the Correlation-based Feature Selection (CFS) filter for the gene selection task. An improved FA was proposed to overcome FA slow convergence by generating mutable size solutions for the firefly population. In addition, a composite position update strategy was designed for the mutable size solutions. The proposed strategy was to balance FA exploration and exploitation in order to address the local optima problem. The proposed hybrid algorithm known as CFS-Mutable Composite Firefly Algorithm (CFS-MCFA) was evaluated on cancer microarray data for biomarker selection along with the deployment of Support Vector Machine (SVM) as the classifier. Evaluation was performed based on two metrics: classification accuracy and size of feature set. The results showed that the CFS-MCFA-SVM algorithm outperforms benchmark methods in terms of classification accuracy and genes subset size. In particular, 100 percent accuracy was achieved on all four datasets and with only a few biomarkers (between one and four). This result indicates that the proposed algorithm is one of the competitive alternatives in feature selection, which later contributes to the analysis of microarray data.

Universiti Utara Malaysia

**Keywords:** Data classification, Firefly algorithm, Gene selection, Microarray data.

## **Acknowledgement**

Alhamdulillah... First of all, I am grateful to Allah (S.W.T) for granting me an opportunity to start this journey and blessing me throughout the studies and helping me to complete my thesis.

I also thankful to my supervisor, Assoc. Prof. Dr. Yuhanis Yusof for her precious guidance and inspiration which assisted me during this study. Supervision is supervision which I could realize from you. It is a blessing to being a student under your supervision.

I also would like to appreciate my little loving son Ammar for his realizing baby mind and express my gratitude to my husband for his sacrifice and motivations. Last but not least, I am thankful to my parents and for everyone who were praying for me to complete my research successfully.



## Table of Contents

Permission to Use .....	i
Abstrak .....	ii
Abstract .....	iii
Acknowledgement .....	iv
Table of Contents .....	v
List of Tables .....	ix
List of Figures .....	xi
List of Appendices .....	xiii
List of Abbreviations .....	xiv
<b>CHAPTER ONE INTRODUCTION .....</b>	<b>1</b>
1.1 Background Study .....	1
1.2 Cancer Classification .....	3
1.3 Problem Statement .....	5
1.4 Research Questions .....	12
1.5 Objectives of Study .....	12
1.6 Scope of Study .....	13
1.7 Operational Definition .....	14
1.8 Significance of Study .....	14
1.9 Thesis Outline .....	15
<b>CHAPTER TWO LITERATURE REVIEW .....</b>	<b>16</b>
2.1 Introduction .....	16
2.2 DNA Microarray .....	17
2.3 Machine Learning Classifiers .....	21
2.3.1 Supervised Machine Learning Algorithms .....	21
2.3.2 Unsupervised Machine Learning Algorithms .....	23
2.4 Feature Selection .....	24
2.4.1 Filters .....	26
2.4.2 Wrappers .....	31
2.4.3 Hybrid Methods .....	36
2.4.4 Gene Selection Approaches on Benchmark Cancer Microarray	

Datasets .....	47
2.4.5 Correlation-based Feature Selection .....	50
2.4.6 Gene Selection Approaches with Support Vector Machine .....	52
2.4.7 Optimization Algorithms .....	53
2.4.7.1 Feature Selection using Meta-heuristic Optimization Algorithms .....	54
2.4.7.2 Gene Selection using Meta-heuristic Optimization Algorithms .....	57
2.5 Firefly Algorithm .....	58
2.5.1 Population Initialization .....	59
2.5.2 Position Update .....	62
2.5.3 Population Reinitialization .....	64
2.5.4 Feature Selection using Firefly Algorithm .....	65
2.5.5 Gene Selection using Firefly Algorithm .....	68
2.6 Summary .....	69
<b>CHAPTER THREE METHODOLOGY .....</b>	<b>71</b>
3.1 Introduction .....	71
3.2 Data Acquisition .....	72
3.2.1 Data Collection .....	72
3.2.2 Data Preprocessing .....	76
3.2.2.1 Normalization .....	77
3.2.2.2 Filtering .....	79
3.3 Algorithm Design .....	80
3.3.1 Parameter Setting .....	80
3.3.2 Population Initialization .....	81
3.3.3 Informative Gene Selection .....	81
3.3.3.1 Objective Function Evaluation .....	81
3.3.3.2 Position Update .....	82
3.3.3.3 Population Reinitialization .....	82
3.3.3.4 Population Refinement .....	83
3.4 Classification .....	83
3.5 Evaluation .....	83
3.6 Summary .....	85

<b>CHAPTER FOUR MUTABLE COMPOSITE FIREFLY ALGORITHM .....</b>	<b>86</b>
4.1 Introduction .....	86
4.2 Non-Fixed Size Firefly Population Generation .....	86
4.3 Composite Firefly Position Update .....	87
4.3.1 Integrative Position Update .....	88
4.3.2 Discriminative Position Update .....	88
4.4 Firefly Population Reinitialization .....	90
4.5 Firefly Population Refinement .....	91
4.6 Mutable Composite Firefly Algorithm for Gene Selection .....	92
4.7 Summary .....	96
<b>CHAPTER FIVE RESULTS AND DISCUSSION .....</b>	<b>97</b>
5.1 Introduction .....	97
5.2 Evaluation .....	97
5.2.1 Comparison between MCFA and Preliminary Methods .....	97
5.2.1.1 Cancer Classification using SVM .....	98
5.2.1.2 Cancer Classification using CFS Filtering and SVM .....	99
5.2.1.3 Cancer Classification using Firefly Algorithm and SVM ..	100
5.2.1.4 Cancer Classification using CFS Filtering with Firefly Algorithm and SVM .....	102
5.2.1.5 Cancer Classification using CFS Filtering with Mutable Composite Firefly Algorithm and SVM .....	104
5.2.1.6 Performance Evaluation between CFS-MCFA-SVM and Preliminary Methods .....	105
5.2.2 Comparison between MCFA and Benchmark Methods .....	109
5.2.2.1 Classification Performance: CFS-MCFA-SVM .....	110
5.2.2.1.1 Performance of CFS-MCFA-SVM with Co-ABC Settings .....	110
5.2.2.1.2 Performance of CFS-MCFA-SVM with FFF-SVM Settings .....	112
5.2.2.1.3 Performance of CFS-MCFA-SVM with rMRMR- MBA Settings .....	114
5.2.2.2 Comparison between CFS-MCFA-SVM and Co-ABC .....	116

5.2.2.2.1 Performance Evaluation between CFS-MCFA-SVM and Co-ABC .....	117
5.2.2.2.2 Paired T-test between Co-ABC-SVM and CFS-MCFA-SVM .....	120
5.2.2.2.3 Convergence Behaviour between CFS-MCFA-SVM and Co-ABC-SVM .....	122
5.2.2.3 Performance Evaluation between CFS-MCFA-SVM and FFF-SVM .....	124
5.2.2.4 Performance Evaluation between CFS-MCFA-SVM and rMRMR-MBA .....	127
<b>CHAPTER SIX CONCLUSIONS AND FUTURE WORKS .....</b>	<b>129</b>
6.1 Conclusions .....	129
6.2 Contributions .....	130
6.3 Limitations of the Study .....	131
6.4 Future Works .....	131
<b>REFERENCES .....</b>	<b>133</b>
<b>APPENDICES .....</b>	<b>157</b>

## List of Tables

Table 2.1	Summary of Gene Selection Studies for Cancer Classification using Filter .....	29
Table 2.2	Summary of Gene Selection Studies for Cancer Classification using Wrapper .....	33
Table 2.3	Summary of Gene Selection Studies for Cancer Classification using Hybrid Method .....	41
Table 2.4	Comparison of Gene Selection Studies on Benchmark Cancer Microarray Datasets .....	48
Table 2.5	Performance of Hybrid Gene Selection Studies on Benchmark Datasets: CFS vs. mRMR .....	51
Table 2.6	Gene Selection Studies for Cancer Classification using Support Vector Machine .....	52
Table 2.7	Summary of Feature Selection Studies using Meta-heuristic Algorithms .....	56
Table 2.8	Summary of Gene Selection Studies for Cancer Classification using Meta-heuristic Algorithms .....	57
Table 2.9	Population Generation in Gene Selection Studies using Firefly Algorithm .....	60
Table 2.10	Position Update in Gene Selection Studies using Firefly Algorithm ....	64
Table 2.11	Summary of Feature Selection Studies using Firefly Algorithm .....	67
Table 2.12	Summary of Gene Selection Studies for Cancer Classification using Firefly Algorithm .....	68
Table 3.1	Elaboration of Research Design .....	74
Table 3.2	Dataset Description .....	75
Table 3.3	Normalized Filtered Dataset .....	80
Table 5.1	Classification Performance obtained on Normalized Microarray Datasets using SVM .....	98
Table 5.2	Classification Performance obtained on Normalized CFS Filtered Microarray Datasets using SVM .....	100
Table 5.3	Classification Performance obtained on Normalized Microarray Datasets using FA and SVM .....	102

Table 5.4 Classification Performance obtained on Normalized CFS Filtered Microarray Datasets using FA and SVM .....	103
Table 5.5 Classification Performance obtained on Normalized CFS Filtered Microarray Datasets using MCFA and SVM .....	104
Table 5.6 Classification Performance Evaluation between CFS-MCFA-SVM and Preliminary Methods .....	106
Table 5.7 Parameter Settings in Benchmark Approaches .....	110
Table 5.8 Classification Performance of CFS-MCFA-SVM using Co-ABC Parameter Settings .....	111
Table 5.9 Informative Genes obtained using CFS-MCFA-SVM with Co-ABC Settings .....	112
Table 5.10 Classification Performance of CFS-MCFA-SVM using FFF-SVM Parameter Settings .....	113
Table 5.11 Informative Genes obtained using CFS-MCFA-SVM with FFF-SVM Settings .....	114
Table 5.12 Classification Performance of CFS-MCFA-SVM using rMRMR-MBA Parameter Settings .....	115
Table 5.13 Informative Genes obtained using CFS-MCFA-SVM with rMRMR-MBA Settings .....	116
Table 5.14 Classification Performance obtained on Normalized CFS Filtered Microarray Datasets using ABC and SVM: Co-ABC-SVM .....	118
Table 5.15 Informative Genes obtained using Co-ABC-SVM .....	118
Table 5.16 Classification Performance Evaluation between CFS-MCFA-SVM and Co-ABC-SVM .....	119
Table 5.17 Classification Performance Evaluation between Benchmark Methods.	121
Table 5.18 P-values between Co-ABC-SVM and CFS-MCFA-SVM .....	121
Table 5.19 Classification Performance of CFS-MCFA-SVM using FFF-SVM Parameter Settings with LOOCV .....	125
Table 5.20 Informative Genes obtained using CFS-MCFA-SVM with FFF-SVM Settings with LOOCV .....	126
Table 5.21 Classification Performance Evaluation between CFS-MCFA-SVM and FFF-SVM .....	126
Table 5.22 Classification Performance Evaluation between CFS-MCFA-SVM and rMRMR-MBA .....	128

## List of Figures

Figure 2.1: Microarray formation .....	17
Figure 2.2: Sample microarray dataset .....	18
Figure 2.3: Taxonomy of optimization algorithms .....	54
Figure 2.4: Pseudo code of Firefly Algorithm .....	59
Figure 3.1: Phases in methodology .....	72
Figure 3.2: Flow of data preprocessing .....	77
Figure 3.3: An example of cancer microarray dataset for Leukemia3 .....	78
Figure 3.4: An example of normalized microarray dataset for Leukemia3 .....	78
Figure 3.5: An example of normalized filtered microarray dataset for Leukemia3....	79
Figure 3.6: Confusion matrix .....	84
Figure 4.1: Population representation with mutable firefly solutions.....	87
Figure 4.2: Pseudo code for mutable firefly generation method .....	87
Figure 4.3: Integrative position update in less bright firefly.....	88
Figure 4.4: Discriminative position update in less bright firefly .....	89
Figure 4.5: Pseudo code for firefly population reinitialization method.....	91
Figure 4.6: Process of firefly population refinement .....	92
Figure 4.7: Flow chart of Mutable Composite Firefly Algorithm .....	93
Figure 4.8: Mutable Composite Firefly Algorithm .....	95
Figure 5.1: Convergence behaviour between CFS-MCFA-SVM and CFS-FA-SVM for Colon dataset .....	107
Figure 5.2: Convergence behaviour between CFS-MCFA-SVM and CFS-FA-SVM for Leukemia2 dataset .....	107
Figure 5.3: Convergence behaviour between CFS-MCFA-SVM and CFS-FA-SVM for Leukemia3 dataset .....	108
Figure 5.4: Convergence behaviour between CFS-MCFA-SVM and CFS-FA-SVM for SRBCT dataset .....	108
Figure 5.5: Convergence behaviour between CFS-MCFA-SVM and Co-ABC-SVM for Colon dataset .....	123
Figure 5.6: Convergence behaviour between CFS-MCFA-SVM and Co-ABC-SVM for Leukemia2 dataset .....	123

Figure 5.7: Convergence behaviour between CFS-MCFA-SVM and Co-ABC-SVM for Leukemia3 dataset .....	124
Figure 5.8: Convergence behaviour between CFS-MCFA-SVM and Co-ABC-SVM for SRBCT dataset .....	124



## **List of Appendices**

Appendix A Classification Results: SVM vs. CFS-SVM .....	157
Appendix B Classification Results: FA-SVM .....	158
Appendix C Classification Results: CFS-FA-SVM .....	160
Appendix D Classification Results: CFS-MCFA-SVM .....	161
Appendix E Classification Results: Co-ABC-SVM .....	166
Appendix F Publications during the Study .....	168



## **List of Abbreviations**

ABC	-	Artificial Bee Colony
ACO	-	Ant Colony Optimization
AEN-CMI	-	Adaptive Elastic Net-Conditional Mutual Information
AHSGS	-	Adaptive Harmony Search Gene Selection
ALL	-	Acute Lymphoblastic Leukemia
AML	-	Acute Myeloid Leukemia
ANN	-	Artificial Neural Network
AntRSAR	-	Ant Colony Based Reduct
AP-AMBFA	-	Affinity Propagation-Adaptive Modified Binary Firefly Algorithm
BA	-	Bat Algorithm
BBA	-	Binary Bat Algorithm
BBBO	-	Binary Biogeography-Based Optimization
BBBOFS	-	Binary Biogeography-Based Optimization Feature Selection
BCWD	-	Breast Cancer Wisconsin datasets in Diagnostic
BCWO	-	Breast Cancer Wisconsin datasets in Original
BeeRSAR	-	Bee Colony Based Reduct
BFA	-	Binary Firefly Algorithm
BFS	-	Best First Search
BL	-	Burkitt's Lymphoma
BSFLA	-	Binary Shuffled Frog Leaping Algorithm
BSO	-	Brain Storm Optimization
CDA	-	Chaotic Dragonfly Algorithm
cDNA	-	complementary Deoxyribo Nucleic Acid
CFS	-	Correlation-based Feature Selection
CLL	-	Chronic Lymphoblastic Leukemia
CMIM	-	Conditional Mutual Information Maximization
CML	-	Chronic Myeloid Leukemia
CNS	-	Central Nervous System
COA	-	Cuckoo Optimization Algorithm
Co-ABC	-	Correlation Artificial Bee Colony

CoFA	-	Composite Firefly Algorithm
CPU	-	Central Processing Unit
CSO	-	Chicken Swarm Optimization
CT	-	Computerized Axial Tomography
CV	-	Cross Validation
DBSCAN	-	Density-Based Spatial Clustering of Applications with Noise
DNA	-	Deoxyribo Nucleic Acid
DNN	-	Deep Neural Network
DT	-	Decision Tree
Duke	-	Duke Breast Cancer Gene dataset
ELM	-	Extreme Learning Machine
ESA	-	Elephant Search Algorithm
EWS	-	Ewing's Sarcoma
FA	-	Firefly Algorithm
FA-FS	-	Firefly Algorithm based Feature Selection
FCBF	-	Fast Correlation-Based Filter
FF	-	Firefly
FFA	-	Firefly Algorithm
FFF	-	f-score-Firefly
FIFS	-	Firefly Inspired Feature Selection
FMI	-	Fuzzy Mutual Information
FP	-	False Positive
FSVM	-	Fuzzy Support Vector Machine
GA	-	Genetic Algorithm
GALA	-	GA and Learning Automata
GB	-	Gigabyte
GBC	-	Genetic Bee Colony
GenRSAR	-	Genetic Based Reduct
GSP	-	Gene Selection Programming
GWO	-	Grey Wolf Optimizer
HCC	-	Hepatocellular Carcinoma
HS	-	Harmony Search

IB1	-	Instance Based
iBPSO	-	improved-Binary Particle Swarm Optimization
IG	-	Information Gain
JMI	-	Joint Mutual Information
JNMIF	-	Joe's Normalized Mutual Information-based Filter
KC	-	Kendall Correlation
KNN	-	K-Nearest Neighbor
LFSDL	-	L1-regulated Feature Selection using Deep Learning
LIBSVM	-	Library for Support Vector Machines
LOOCV	-	Leave-One-Out Cross Validation
MBA	-	Modified Bat Algorithm
MBF	-	Markov Blanket Filter
MCFA	-	Mutable Composite Firefly Algorithm
MI	-	Mutual Information
MIFS	-	Mutual Information based Feature Selection
MIMAGA	-	Mutual Information Maximization-Adaptive Genetic Algorithm
ML	-	Machine Learning
MOBBA-LS	-	Multi-Objective Binary Bat Algorithm with Local Search
MRI	-	Magnetic Resonance Imaging
mRMR	-	minimum Redundancy Maximum Relevance
NB	-	Naïve Bayes
NB	-	Neuroblastoma
NCI	-	National Cancer Institute
ORL	-	Olivetti Research Lab
PC	-	Personal Computer
PCA	-	Principal Component Analysis
PET	-	Positron Emission Tomography
PSO	-	Particle Swarm Optimization
PSO-RSAR	-	Particle Swarm Based Reduct
RAM	-	Random Access Memory
RF	-	Random Forest
RFA	-	Recursive Firefly Algorithm

rMRMR	-	robust Minimum Redundancy Maximum Relevance
RMS	-	Rhabdomyosarcoma
ROC	-	Receiver Operating Characteristic
RSAR	-	Rough Set based Attribute Reduction
SC	-	Spearman's Correlation
SMOTE	-	Synthetic Minority Over-sampling Technique
SNR	-	Signal-to-Noise Ratio
SRBCT	-	Small Round Blue Cell Tumor
SSO	-	Simplified Swarm Optimization
SVM	-	Support Vector Machine
TP	-	True Positive
TRIZ	-	Teoriya Resheniya Izobretatelskikh Zadatch
WDBC	-	Wisconsin Diagnosis Breast Cancer
WEKA	-	Waikato Environment for Knowledge Analysis
WHO	-	World Health Organization
WOA	-	Whale Optimization Algorithm



# **CHAPTER ONE**

## **INTRODUCTION**

### **1.1 Background Study**

Meta-heuristic optimization algorithms (Al-Betar et al., 2020; Almugren & Alshamlan, 2019c; Alshamlan, 2018; Dash, 2018; Elyasigomari et al., 2017) have shown competing performance in gene selection for cancer classification. Firefly Algorithm (FA) is a popular meta-heuristic optimization algorithm with two special properties: automatically subdivision and the ability for multimodality which makes FA efficient in optimization and classification problems (Yang & He, 2013). Further FA is capable of balancing between exploration and exploitation (Yang & He, 2013). However, FA suffers from two major drawbacks: slow convergence and trapped into local optimums (Dif & Elberrichi, 2019; Mazen et al., 2016; Sharma & Tyagi, 2022). Hence, some modifications are required to overcome these issues prior to the application in feature selection on big datasets such as cancer microarray. Thus, a new variant of FA with mutable and composite properties is introduced in this study. Mutable property represents the dynamic nature of the fireflies or solutions in the population in contrast to fixed size solutions as reported in existing studies (Almugren & Alshamlan, 2019c; Emami & Pakzad, 2019; Peng et al., 2020). Further the term composite denotes the composite strategy used for position update function in the proposed study.

The convergence property of a swarm algorithm can be represented as the capability of finding the global best solution quickly. For instance, assume there are two algorithms: algorithm1 and algorithm2 and maximum iteration is equal to 100; and for

## References

- Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020, April). Machine Learning Supervised Algorithms of Gene Selection: A Review. *Machine Learning*, 62(03), 233-244.
- Agarwal, V., & Bhanot, S. (2015, August). Firefly inspired feature selection for face recognition. In *2015 eighth international conference on contemporary computing (IC3)* (pp. 257-262). IEEE. <https://doi.org/10.1109/IC3.2015.7346689>
- Aghdam, M. H., Ghasem-Aghaee, N., & Basiri, M. E. (2009). Text feature selection using ant colony optimization. *Expert systems with applications*, 36(3), 6843-6853. <https://doi.org/10.1016/j.eswa.2008.08.022>
- Aha, D. W., Kibler, D., & Albert, M. K. (1991, January). Instance-based learning algorithms. *Mach Learn*, 6, 37-66. <https://doi.org/10.1007/BF00153759>
- Al-Ani, A. (2005). Feature subset selection using ant colony optimization. *International journal of computational intelligence (IJCI)*, 2(1), 53-58.
- Alanni, R., Hou, J., Azzawi, H., & Xiang, Y. (2019). A novel gene selection algorithm for cancer classification using microarray datasets. *BMC Medical Genomics*, 12(1), 1-12. <https://doi.org/10.1186/s12920-018-0447-6>
- Alba, E., Garcia-Nieto, J., Jourdan, L., & Talbi, E. G. (2007, September). Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms. In *2007 IEEE congress on evolutionary computation* (pp. 284-290). IEEE. <https://doi.org/10.1109/CEC.2007.4424483>
- Al-Batah, M., Zaqaibeh, B., Alomari, S. A., & Alzboon, M. S. (2019). Gene Microarray Cancer Classification using Correlation Based Feature Selection

Algorithm and Rules Classifiers. *International Journal of Online & Biomedical Engineering (iJOE)*, 15(8), 62-73.  
<https://doi.org/10.3991/ijoe.v15i08.10617>

Al-Betar, M. A., Alomari, O. A., & Abu-Romman, S. M. (2020). A TRIZ-inspired bat algorithm for gene selection in cancer classification. *Genomics*, 112(1), 114-126. <https://doi.org/10.1016/j.ygeno.2019.09.015>

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., ... Staudt, L. M. (2000, February 3). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769), 503-511. <https://doi.org/10.1038/35000501>

Almugren, N., & Alshamlan, H. (2019a, June). A survey on hybrid feature selection methods in microarray gene expression data for cancer classification. *IEEE Access*, 7, 78533-78548. <https://doi.org/10.1109/ACCESS.2019.2922987>

Almugren, N., & Alshamlan, H. (2019b, July). FF-SVM: New FireFly-based Gene Selection Algorithm for Microarray Cancer Classification. In *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1-6). <https://doi.org/10.1109/CIBCB.2019.8791236>

Almugren, N., & Alshamlan, H. M. (2019c). New bio-marker gene discovery algorithms for cancer gene expression profile. *IEEEAccess*, 7, 136907-136913. <https://doi.org/10.1109/ACCESS.2019.2942413>

Alomari, O. A., Khader, A. T., Al-Betar, M. A., & Abualigah, L. M. (2017). MRMR BA: a hybrid gene selection algorithm for cancer classification. *J Theor Appl Inf Technol*, 95(12), 2610-2618.

Alon, U., Barkai, N., Notterman, D., Gish, K., Ybarra, S., Mack, D., & Levine, A. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *PNAS*. 96(12), 6745-6750. <https://doi.org/10.1073/pnas.96.12.6745>

Alonso-González, C. J., Moro-Sancho, Q. I., Simon-Hurtado, A., & Varela-Arrabal, R. (2012). Microarray gene expression classification with few genes: Criteria to combine attribute selection and classification methods. *Expert Systems with Applications*, 39(8), 7270-7280. <https://doi.org/10.1016/j.eswa.2012.01.096>

Al Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9), 735-739.

Alshamlan, H., Badr, G., & Alohal, Y. (2014). A comparative study of cancer classification methods using microarray gene expression profile. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (pp. 389-398). Springer, Singapore. [https://doi.org/10.1007/978-981-4585-18-7\\_44](https://doi.org/10.1007/978-981-4585-18-7_44)

Universiti Utara Malaysia  
Alshamlan, H., Badr, G., & Alohal, Y. (2015a). mRMR-ABC: A hybrid gene selection algorithm for cancer classification using microarray gene expression profiling. *BioMed research international*, 2015. <https://doi.org/10.1155/2015/604910>

Alshamlan, H. M. (2018). Co-ABC: Correlation artificial bee colony algorithm for biomarker gene discovery using gene expression profile. *Saudi Journal of Biological Sciences*, 25(5), 895-903. <https://doi.org/10.1016/j.sjbs.2017.12.012>

Alshamlan, H. M., Badr, G. H., & Alohal, Y. A. (2015b). Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification. *Computational biology and chemistry*, 56, 49-60. <https://doi.org/10.1016/j.compbiochem.2015.03.001>

AL-Taie, F. A. Y., & Qasim, O. S. (2020). Classification of diseases using a hybrid fuzzy mutual information technique with binary bat algorithm. *Ann Proteom Bioinform*, 4, 001-005. <https://doi.org/10.29328/journal.apb.1001009>

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., & Korsmeyer, S. J. (2002, January). Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet*. 30(1), 41-47. <https://doi.org/10.1038/ng765>

Aziz, R., Verma, C. K., & Srivastava, N. (2016). A fuzzy based feature selection from independent component subspace for machine learning classification of microarray data. *Genomics Data*, 8, 4-15. <https://doi.org/10.1016/j.gdata.2016.02.012>

Banati, H., & Bajaj, M. (2011). Fire fly based feature selection approach. *International Journal of Computer Science Issues (IJCSI)*, 8(4), 473-480.

Bayes, T. (1763, January 1). *LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S* (Vol. 53, pp. 370-418). Philosophical transactions of the Royal Society of London. <https://doi.org/10.1098/rstl.1763.0053>

Beer, D. G., Kardia, S. L., Huang, C.-C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., Lizyness, M. L., Kuick, R., Hayasaka, S., Taylor, J. M. G., Iannettoni, M. D., Orringer, M. B., & Hanash, S. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*, 8(8), 816-824. <https://doi.org/10.1038/nm733>

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., & Meyerson, M. (2001, November 20). Classification of human lung carcinomas by mRNA

expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences*, 98(24), 13790-13795. <https://doi.org/10.1073/pnas.191502998>

Bidar, M., & Kanan, H. R. (2013, October). Jumper firefly algorithm. In *ICCKE 2013* (pp. 267-271). IEEE. <https://doi.org/10.1109/ICCKE.2013.6682839>

Bir-Jmel, A., Douiri, S. M., & Elberoussi, S. (2019). Gene selection via a new hybrid ant colony optimization algorithm for cancer classification in high-dimensional data. *Computational and mathematical methods in medicine*, 2019, 1-20. <https://doi.org/10.1155/2019/7828590>

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014, October 20). A review of microarray datasets and applied feature selection methods. *Information Sciences*, 282, 111-135. <https://doi.org/10.1016/j.ins.2014.05.042>

Botstein, D., Brown, P., & Staudt, L. (2000). Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769), 503-511. <https://doi.org/10.1038/35000501>

Cancer Treatment Centers of America. (2021, April 2). About leukemia. <https://www.cancercenter.com/cancer-types/leukemia/about>

Chandrashekhar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>

Cheng, J., Greiner, R., Kelly, J., Bell, D., & Liu, W. (2002, May). Learning Bayesian networks from data: An information-theory based approach. *Artificial intelligence*, 137(1-2), 43-90. [https://doi.org/10.1016/S0004-3702\(02\)00191-1](https://doi.org/10.1016/S0004-3702(02)00191-1)

Cheng, S., Shi, Y., Qin, Q., Ting, T. O., & Bai, R. (2014a, July 6-11). Maintaining population diversity in brain storm optimization algorithm. In *2014 IEEE*

*Congress on Evolutionary Computation (CEC)* (pp. 3230-3237). IEEE.  
<https://doi.org/10.1109/CEC.2014.6900255>

Cheng, S., Shi, Y., Qin, Q., Zhang, Q., & Bai, R. (2014b). Population diversity maintenance in brain storm optimization algorithm. *Journal of Artificial Intelligence and Soft Computing Research*, 4(2), 83-97.  
<https://doi.org/10.1515/jaiscr-2015-0001>

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1), 113-126.  
[https://doi.org/10.1016/S0893-6080\(03\)00169-2](https://doi.org/10.1016/S0893-6080(03)00169-2)

Cortes, C., & Vapnik, V. (1995, September). Support-vector networks. *Mach Learn*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>

Costa, J. C., Tavares, R., & Rosa, A. (1999, October 12-15). An experimental study on dynamic random variation of population size. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* (Vol. 1, pp. 607-612). IEEE.  
<https://doi.org/10.1109/ICSMC.1999.814161>

Cover, T., & Hart, P. (1967, January). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.  
<https://doi.org/10.1109/TIT.1967.1053964>

Das, S., Abraham, A., & Konar, A. (2009). *Metaheuristic clustering. Studies in computational intelligence* (1st ed., Vol. 178). Springer.

Dash, R. (2021). An Adaptive Harmony Search Approach for Gene Selection and Classification of High Dimensional Medical Data. *Journal of King Saud University – Computer and Information Sciences*, 33(2), 195-207.  
<https://doi.org/10.1016/j.jksuci.2018.02.013>

Dash, R., Dash, R., & Rautray, R. (2019). An evolutionary framework based microarray gene selection and classification approach using binary shuffled frog leaping algorithm, *Journal of King Saud University – Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2019.04.002>

Dashtban, M., & Balafar, M. (2017, March). Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics*, 109(2), 91-107. <https://doi.org/10.1016/j.ygeno.2017.01.004>

Dashtban, M., Balafar, M., & Suravajhala, P. (2018). Gene selection for tumor classification using a novel bio-inspired multi-objective approach. *Genomics*, 110(1), 10-17. <https://doi.org/10.1016/j.ygeno.2017.07.010>

Deb, S., Fong, S., & Tian, Z. (2015, October). Elephant search algorithm for optimization problems. In *2015 Tenth International Conference on Digital Information Management (ICDIM)* (pp. 249-255). IEEE. <https://doi.org/10.1109/ICDIM.2015.7381893>

Dechter, R., & Pearl, J. (1985). Generalized best-first search strategies and the optimality of A\*. *Journal of the ACM (JACM)*, 32(3), 505-536. <https://doi.org/10.1145/3828.3830>

Dif, N., & Elberrichi, Z. (2019). An Enhanced Recursive Firefly Algorithm for Informative Gene Selection. *International Journal of Swarm Intelligence Research (IJSIR)*, 10(2), 21-33. <https://doi.org/10.4018/IJSIR.2019040102>

Dokeroglu, T., Sevinc, E., Kucukyilmaz, T., & Cosar, A. (2019). A survey on new generation metaheuristic algorithms. *Computers & Industrial Engineering*, 137, 106040. <https://doi.org/10.1016/j.cie.2019.106040>

Dorigo, M., Maniezzo, V., & Colomi, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 26(1), 29-41. <https://doi.org/10.1109/3477.484436>

Dorj, U. O., Lee, K. K., Choi, J. Y., & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, 77(8), 9909-9924. <https://doi.org/10.1007/s11042-018-5714-1>

Durga, T.S., & Yasaswini, V. (2016). An Enhancement for the optimization of feature selection to perform classification Using Meta Heuristic Algorithms. *International Journal of Latest Engineering Research and Applications (IJLERA)*, 1(9), 64-70.

El-Abd, M. (2016, July 24-29). Brain storm optimization algorithm with re-initialized ideas and adaptive step size. In *2016 IEEE Congress on Evolutionary Computation (CEC)* (pp. 2682-2686). IEEE. <https://doi.org/10.1109/CEC.2016.7744125>

El Akadi, A., Amine, A., El Ouardighi, A., & Aboutajdine, D. (2011). A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. *Knowledge and Information Systems*, 26(3), 487-500. <https://doi.org/10.1007/s10115-010-0288-x>

Elyasigomari, V., Lee, D. A., Screen, H. R. C., & Shaheed, M. H. (2017). Development of a two-stage gene selection method that incorporates a novel hybrid approach using the cuckoo optimization algorithm and harmony search for cancer classification. *Journal of Biomedical Informatics*, 67, 11-20. <https://doi.org/10.1016/j.jbi.2017.01.016>

Emami, N., & Pakzad, A. (2019). A New Knowledge-Based System for Diagnosis of Breast Cancer by a combination of the Affinity Propagation and Firefly Algorithms. *Journal of AI and Data Mining*, 7(1), 59-68. <https://doi.org/10.22044/JADM.2018.6489.1763>

Emary, E., Zawbaa, H. M., Ghany, K. K. A., Hassanien, A. E., & Parv, B. (2015, September). Firefly optimization algorithm for feature selection. In *Proceedings of the 7th Balkan Conference on Informatics Conference* (pp. 1-7). <https://doi.org/10.1145/2801081.2801091>

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

Fajriyah, R. (2021). Paper review: An overview on microarray technologies. *Bulletin of Applied Mathematics and Mathematics Education*, 1(1), 21-30. <https://doi.org/10.12928/bamme.v1i1.3854>

Fathi, E., Mesbah-Namin, S. A., & Farahzadi, R. (2014). Biomarkers in medicine: an overview. *Journal of Advances in Medicine and Medical Research*, 4(8), 1701-1718. <https://doi.org/10.9734/BJMMR/2014/6917>

Fernandes, K., Cardoso, J. S., & Fernandes, J. (2017, June). Transfer learning with partial observability applied to cervical cancer screening. In *Iberian conference on pattern recognition and image analysis* (Vol. 10255, pp. 243-250). Springer, Cham. [https://doi.org/10.1007/978-3-319-58838-4\\_27](https://doi.org/10.1007/978-3-319-58838-4_27)

Fong, S., Biuk-Aghai, R. P., & Millham, R. C. (2018, February). Swarm search methods in Weka for data mining. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (pp. 122-127). <https://doi.org/10.1145/3195106.3195167>

Fu, S., & Desmarais, M. C. (2010, June). Markov blanket based feature selection: a review of past decade. In *Proceedings of the world congress on engineering* (Vol. 1, pp. 321-328). Newswood Ltd.

Gao, L., Ye, M., Lu, X., & Huang, D. (2017). Hybrid method based on information gain and support vector machine for gene selection in cancer classification.

*Genomics, proteomics & bioinformatics*, 15(6), 389-395.  
<https://doi.org/10.1016/j.gpb.2017.08.002>

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation*, 76(2), 60-68.  
<https://doi.org/10.1177/003754970107600201>

Ghoniem, R. M. (2020). A novel bio-inspired deep learning approach for liver cancer diagnosis. *Information*, 11(2), 80. <https://doi.org/10.3390/info11020080>

Goldberg, D. E. (1989, June 1). Sizing populations for serial and parallel genetic algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms and Their Applications* (pp. 70-79). Morgan Kaufmann.

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C., & Lander, E. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439), 531-537.  
<https://doi.org/10.1126/science.286.5439.531>

Gordon, G. J., Jensen, R. V., Hsiao, L. L., Gullans, S. R., Blumenstock, J. E., Ramaswamy, S., Richards, W. G., Sugarbaker, D. J., & Bueno, R. (2002, September 1). Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer research*, 62(17), 4963-4967.

Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). Microarray and its applications. *J Pharm Bioallied Sci.*, 4(Suppl 2), S310-S312.  
<https://doi.org/10.4103/0975-7406.100283>

Gunavathi, C., & Premalatha, K. (2014). Performance analysis of genetic algorithm with KNN and SVM for feature selection in tumor classification. *International Journal of Computer and Information Engineering*, 8(8), 1490-1497.

Guo, J., & Tang, S. J. (2009, August 26-27). An improved particle swarm optimization with re-initialization mechanism. In *2009 International Conference on Intelligent Human-Machine Systems and Cybernetics* (Vol. 1, pp. 437-441). IEEE. <https://doi.org/10.1109/IHMSC.2009.117>

Guo, Q. M. (2003). DNA microarray and cancer. *Current opinion in oncology*, 15(1), 36-43.

Guyon, I., Boser, B., & Vapnik, V. (1993). Automatic capacity tuning of very large VC-dimension classifiers. In *Advances in neural information processing systems* (pp. 147-155).

Hall, M. A. (1999). Correlation-based feature selection for machine learning [Doctoral dissertation, University of Waikato]. University of Waikato. <https://www.cs.waikato.ac.nz/~mhall/thesis.pdf>

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417-441. <https://doi.org/10.1037/h0071325>

Inza, I., Larrañaga, P., Blanco, R., & Cerrolaza, A. J. (2004). Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial intelligence in medicine*, 31(2), 91-103. <https://doi.org/10.1016/j.artmed.2004.01.007>

Jabbar, S. F. (2019). A classification model on tumor cancer disease based mutual information and firefly algorithm. *Periodicals of Engineering and Natural Sciences*, 7(3), 1152-1162. <https://doi.org/10.21533/pen.v7i3.656>

Jain, I., Jain, V. K., & Jain, R. (2018). Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. *Applied Soft Computing*, 62, 203-215. <https://doi.org/10.1016/j.asoc.2017.09.038>

Johnson, S. C. (1967, September). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241-254. <https://doi.org/10.1007/BF02289588>

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed., Vol. XXIX, pp. 488). Springer. <https://doi.org/10.1007/b98835>

Kang, J. (2015). Clinical implications of microarray in cancer medicine. *International Journal of Cancer Research*, 11(4), 150-158. <https://doi.org/10.3923/ijcr.2015.150.158>

Kar, S., Sharma, K. D., & Maitra, M. (2015). Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. *Expert Systems with Applications*, 42(1), 612-627. <https://doi.org/10.1016/j.eswa.2014.08.014>

Karley, D., Gupta, D., & Tiwari, A. (2011, January). Biomarkers: The future of medical science to detect cancer. *J Mol Biomark Diagn*, 2(5), 1-7. <https://doi.org/10.4172/2155-9929.1000118>

Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., & Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 7(6), 673-679. <https://doi.org/10.1038/89044>

Kim, T. K. (2015). T test as a parametric statistic. *Korean journal of anesthesiology*, 68(6), 540-546. <https://doi.org/10.4097/kjae.2015.68.6.540>

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273-324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.

Koumousis, V. K., & Katsaras, C. P. (2006, February). A saw-tooth genetic algorithm combining the effects of variable population size and reinitialization to enhance performance. *IEEE Transactions on Evolutionary Computation*, 10(1), 19-28. <https://doi.org/10.1109/TEVC.2005.860765>

Lai, C. M., & Huang, H. P. (2021, March). A gene selection algorithm using simplified swarm optimization with multi-filter ensemble technique. *Applied Soft Computing*, 100, 106994. <https://doi.org/10.1016/j.asoc.2020.106994>

Lai, C. M., Yeh, W. C., & Chang, C. Y. (2016). Gene selection using information gain and improved simplified swarm optimization. *Neurocomputing*, 218, 331-338. <https://doi.org/10.1016/j.neucom.2016.08.089>

Leon, S. A., Shapiro, B., Sklaroff, D. M., & Yaros, M. J. (1977, March 1). Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*, 37(3), 646-650.

Leslie, C., Eskin, E., & Noble, W. S. (2001). The spectrum kernel: A string kernel for SVM protein classification. In *Biocomputing 2002* (pp. 564-575). [https://doi.org/10.1142/9789812799623\\_0053](https://doi.org/10.1142/9789812799623_0053)

Liu, H., Li, J., & Wong, L. (2002). A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome informatics*, 13, 51-60. <https://doi.org/10.11234/gi1990.13.51>

Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y., & Gao, Z. (2017). A hybrid feature selection algorithm for gene expression data classification. *Neurocomputing*, 256, 56-62. <https://doi.org/10.1016/j.neucom.2016.07.080>

MacQueen, J. (1967, January 1). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

Marie-Sainte, S. L., & Alalyani, N. (2020). Firefly algorithm based feature selection for Arabic text classification. *Journal of King Saud University-Computer and Information Sciences*, 32(3), 320-328.  
<https://doi.org/10.1016/j.jksuci.2018.06.004>

Mazen, F., AbulSeoud, R. A., & Gody, A. M. (2016). Genetic algorithm and firefly algorithm in a hybrid approach for breast cancer diagnosis. *International Journal of Computer Trends and Technology (IJCTT)*, 32(2), 62-68.

Mazumder, D. H., & Veilumuthu, R. (2018a). An Enhanced Gene Selection Methodology for Effective Microarray Cancer Data Classification. *Int. J. Simulation: Systems, Sci. Technol.*, 19(2).  
<https://doi.org/10.5013/IJSSST.a.19.02.04>

Mazumder, D. H., & Veilumuthu, R. (2018b). Cancer Classification with a Novel Hybrid Feature Selection Technique. *International Journal of Simulation: Systems, Science & Technology*, 19(2).  
<https://doi.org/10.5013/IJSSST.a.19.02.07>

Mazumder, D. H., & Veilumuthu, R. (2019). An enhanced feature selection filter for classification of microarray cancer data. *ETRI Journal*, 41(3), 358-370.  
<https://doi.org/10.4218/etrij.2018-0522>

McCulloch, W. S., & Pitts, W. (1943, December). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133. <https://doi.org/10.1007/BF0247825>

Mostafa Bozorgi, S., & Yazdani, S. (2019, July). IWOA: An improved whale optimization algorithm for optimization problems. *Journal of Computational*

*Design and Engineering*, 6(3), 243-259.  
<https://doi.org/10.1016/j.jcde.2019.02.002>

Moteghaed, N. Y., Maghooli, K., & Garshasbi, M. (2018). Improving classification of cancer and mining biomarkers from gene expression profiles using hybrid optimization algorithms and fuzzy support vector machine. *J Med Signals Sens*, 8(1), 1-11.

Motieghader, H., Najafi, A., Sadeghi, B., & Masoudi-Nejad, A. (2017). A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata. *Informatics in Medicine Unlocked*, 9, 246-254. <https://doi.org/10.1016/j.imu.2017.10.004>

National Cancer Institute. (2015, February 9). What is Cancer?.  
<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>

Nutt, C. L., Mani, D. R., Betensky, R. A., Tamayo, P., Cairncross, J. G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M. E., Batchelor, T. T., Black, P. M., Deimling, A., Pomeroy, S. L., Golub, T. R., & Louis, D. N. (2003, April 1). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer research*, 63(7), 1602-1607.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017, June). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.  
<https://doi.org/10.14445/22312803/IJCTT-V48P126>

Othman, M. S., Kumaran, S. R., & Yusuf, L. M. (2020). Gene Selection Using Hybrid Multi-Objective Cuckoo Search Algorithm with Evolutionary Operators for Cancer Microarray Data. *IEEE Access*, 8, 186348-186361.  
<https://doi.org/10.1109/ACCESS.2020.3029890>

Panda, M. (2020). Elephant search optimization combined with deep neural network for microarray data analysis. *Journal of King Saud University-Computer and Information Sciences*, 32(8), 940-948. <https://doi.org/10.1016/j.jksuci.2017.12.002>

Pashaei, E., & Pashaei, E. (2019, November 28-30). Gene selection using intelligent dynamic genetic algorithm and random forest. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)* (pp. 470-474). IEEE. <https://doi.org/10.23919/ELECO47770.2019.8990557>

Pearl, J. (1984). *Heuristics: intelligent search strategies for computer problem solving*. United States.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572. <https://doi.org/10.1080/14786440109462720>

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8), 1226-1238. <https://doi.org/10.1109/TPAMI.2005.159>

Peng, H., Zhu, W., Deng, C., Yu, K., & Wu, Z. (2020). Composite firefly algorithm for breast cancer recognition. *Concurrency and Computation: Practice and Experience*. <https://doi.org/10.1002/cpe.6032>

Pham, D. T., & Castellani, M. (2009). The Bees Algorithm: Modelling foraging behaviour to solve continuous optimization problems. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical*

*Engineering Science*, 223(12), 2919-2938.  
<https://doi.org/10.1243/09544062JMES1494>

Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y. H., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., ... Golub, T. R. (2002, January 24). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415(6870), 436-442. <https://doi.org/10.1038/415436a>

Quinlan, J. R. (1986, March). Induction of decision trees. *Mach Learn*, 1(1), 81-106.  
<https://doi.org/10.1007/BF00116251>

Ramasubramanian, S. (2020). Role of Microarray in Cancer Biology. *Journal of Complementary Medicine Research*, 11(3), 262-268.  
<https://doi.org/10.5455/jcmr.2020.11.03.33>

Ramaswamy, S., & Golub, T. R. (2002). DNA microarrays in clinical oncology. *Journal of clinical oncology*, 20(7), 1932-1941.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., & Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*, 6(1), 1-6. [https://doi.org/10.1016/S1476-5586\(04\)80047-2](https://doi.org/10.1016/S1476-5586(04)80047-2)

Saengsiri, P., Meesad, P., Wichian, S. N., & Herwig, U. (2010, November). Comparison of hybrid feature selection models on gene expression data. In *2010 Eighth International Conference on ICT and Knowledge Engineering* (pp. 13-18). IEEE. <https://doi.org/10.1109/ICTKE.2010.5692905>

Saeys, Y., Inza, I., & Larriaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507-2517.  
<https://doi.org/10.1093/bioinformatics/btm344>

- Salgotra, R., Singh, U., & Saha, S. (2019). On some improved versions of whale optimization algorithm. *Arabian Journal for Science and Engineering*, 44(11), 9653-9691. <https://doi.org/10.1007/s13369-019-04016-0>
- Santos, M. S., Abreu, P. H., García-Laencina, P. J., Simão, A., & Carvalho, A. (2015, December). A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. *Journal of biomedical informatics*, 58, 49-59. <https://doi.org/10.1016/j.jbi.2015.09.012>
- Sawhney, R., Mathur, P., & Shankar, R. (2018, May). A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis. In *International Conference on Computational Science and Its Applications* (pp. 438-449). Springer, Cham. [https://doi.org/10.1007/978-3-319-95162-1\\_30](https://doi.org/10.1007/978-3-319-95162-1_30)
- Sayed, G. I., Tharwat, A., & Hassanien, A. E. (2019). Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection. *Applied Intelligence*, 49(1), 188-205. <https://doi.org/10.1007/s10489-018-1261-8>
- Sekaj, I., & Oravec, M. (2009, June). Selected population characteristics of fine-grained parallel genetic algorithms with re-initialization. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 945-948). <https://doi.org/10.1145/1543834.1543980>
- Sekaj, I., & Perkacz, J. (2007, September 25-28). Some aspects of parallel genetic algorithms with population re-initialization. In *2007 IEEE Congress on Evolutionary Computation* (pp. 1333-1338). IEEE. <https://doi.org/10.1109/CEC.2007.4424625>
- Sharma, M., & Tyagi, S. (2022, March 29). An Improved Firefly Algorithm Based Cluster Analysis Technique. In *Advances in Data Computing, Communication and Security* (pp. 145-155). Springer, Singapore. [https://doi.org/10.1007/978-981-16-8403-6\\_13](https://doi.org/10.1007/978-981-16-8403-6_13)

Shekar, B. H., & Dagnew, G. (2020). L1-regulated feature selection and classification of microarray cancer data using deep learning. In *Proceedings of 3rd international conference on computer vision and image processing* (pp. 227-242). Springer, Singapore. [https://doi.org/10.1007/978-981-32-9291-8\\_19](https://doi.org/10.1007/978-981-32-9291-8_19)

Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. W., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C., & Golub, T. R. (2002, January 1). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature medicine*, 8(1), 68-74. <https://doi.org/10.1038/nm0102-68>

Shukla, A. K., Singh, P., & Vardhan, M. (2019). DNA gene expression analysis on diffuse large b-cell lymphoma (dlbcl) based on filter selection method with supervised classification method. In *Computational Intelligence in Data Mining* (pp. 783-792). Springer, Singapore. [https://doi.org/10.1007/978-981-10-8055-5\\_69](https://doi.org/10.1007/978-981-10-8055-5_69)

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., & Sellers, W. R. (2002, March). Gene expression correlates of clinical prostate cancer behavior. *Cancer cell*, 1(2), 203-209. [https://doi.org/10.1016/S1535-6108\(02\)00030-2](https://doi.org/10.1016/S1535-6108(02)00030-2)

Singh, P., Shukla, A., & Vardhan, M. (2017, November 23-24). A novel filter approach for efficient selection and small round blue-cell tumor cancer detection using microarray gene expression data. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 827-831). IEEE. <https://doi.org/10.1109/ICICI.2017.8365252>

Sriram, K. B., Larsen, J. E., Yang, I. A., Bowman, R. V., & Fong, K. M. (2011). Genomic medicine in non-small cell lung cancer: Paving the path to

personalized care. *Respirology*, 16(2), 257-263.  
<https://doi.org/10.1111/j.1440-1843.2010.01892.x>

Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J., Scherf, U., Lee, J. K., Reinhold, W. O., Weinstein, J. N., Mesirov, J. P., Lander, E. S., & Golub, T. R. (2001, September 11). Chemosensitivity prediction by transcriptional profiling. *Proceedings of the National Academy of Sciences*, 98(19), 10787-10792. <https://doi.org/10.1073/pnas.191368598>

Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993, July 29). Nuclear feature extraction for breast tumor diagnosis. In *Biomedical image processing and biomedical visualization* (Vol. 1905, pp. 861-870). International Society for Optics and Photonics. <https://doi.org/10.1117/12.148698>

Su, A. I., Welsh, J. B., Sapinoso, L. M., Kern, S. G., Dimitrov, P., Lapp, H., Schultz, P. G., Powell, S. M., Moskaluk, C. A., Frierson H. F., Jr., & Hampton, G. M. (2001, October 15). Molecular classification of human carcinomas by use of gene expression signatures. *Cancer research*, 61(20), 7388-7393.

Subanya, B., & Rajalaxmi, R. R. (2014, February). Feature selection using Artificial Bee Colony for cardiovascular disease classification. In *2014 International Conference on Electronics and Communication Systems (ICECS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ECS.2014.6892729>

Taheri, S., & Mammadov, M. (2013). Learning the naive Bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science*, 23(4), 787-795. <https://doi.org/10.2478/amcs-2013-0059>

Tan, T. Z., Yang, H., Ye, J., Low, J., Choolani, M., Tan, D. S. P., Thiery, J.-P., & Huang, R. Y.-J. (2015). CSIOVDB: a microarray gene expression database of epithelial ovarian cancer subtype. *Oncotarget*, 6(41), 43843-43852. <https://doi.org/10.18632/oncotarget.5983>

Tang, R., Fong, S., Yang, X.-S., & Deb, S. (2012, August). Wolf search algorithm with ephemeral memory. In *Seventh International Conference on Digital Information Management (ICDIM 2012)* (pp. 165-172). IEEE. <https://doi.org/10.1109/ICDIM.2012.6360147>

Van De Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A. M., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., Van Der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., ... Bernards, R. (2002, December 19). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25), 1999-2009. <https://doi.org/10.1056/NEJMoa021967>

Vapnik, V., Golowich, S. E., & Smola, A. J. (1997). Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems 9* (pp. 281-287). MIT Press.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999. <https://doi.org/10.1109/72.788640>

Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. (2002, January 31). Gene expression profiling predicts clinical outcome of breast cancer. *nature*, 415(6871), 530-536. <https://doi.org/10.1038/415530a>

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Comput & Applic*, 24(1), 175-186. <https://doi.org/10.1007/s00521-013-1368-0>

Wang, A., & Gehan, E. A. (2005, April 1). Gene selection for microarray data analysis using principle component analysis. *Stat. Med*, 24(13), 2069-2087. <https://doi.org/10.1002/sim.2082>

Wang, L., Ni, H., Yang, R., Pappu, V., Fenn, M. B., & Pardalos, P. M. (2014). Feature selection based on meta-heuristics for biomedicine. *Optimization Methods and Software*, 29(4), 703-719. <https://doi.org/10.1080/10556788.2013.834900>

Wang, Y., Tetko, I. V., Hall, M. A., Frank, E., Facius, A., Mayer, K. F., & Mewes, H. W. (2005, February). Gene selection from microarray data for cancer classification—a machine learning approach. *Computational biology and chemistry*, 29(1), 37-46. <https://doi.org/10.1016/j.combiolchem.2004.11.001>

Wang, Y., Yang, X.-G., & Lu, Y. (2019). Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information. *Applied Mathematical Modelling*, 71, 286-297. <https://doi.org/10.1016/j.apm.2019.01.044>

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Jr., Marks, J. R., & Nevins, J. R. (2001, September 25). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proceedings of the National Academy of Sciences*, 98(20), 11462-11467. <https://doi.org/10.1073/pnas.201162998>

Weston, J., & Watkins, C. (1999, April). Support vector machines for multi-class pattern recognition. In *Esann* (Vol. 99, pp. 219-224).

Wolberg, W. H., & Mangasarian, O. L. (1990, December 1). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the national academy of sciences*, 87(23), 9193-9196. <https://doi.org/10.1073/pnas.87.23.9193>

World Health Organization. (2018, September 12). Cancer. <https://www.who.int/news-room/fact-sheets/detail/cancer>

Wright, S. (1965, September). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19(3), 395-420. <https://doi.org/10.2307/2406450>

Yadav, J., & Sharma, M. (2013, July). A Review of K-mean Algorithm. *International journal of engineering trends and technology*, 4(7), 2972-2976.

Yang, X.-S. (2010a). *Nature-inspired metaheuristic algorithms*. Luniver press.

Yang, X.-S. (2010b). A new metaheuristic bat-inspired algorithm. In *Nature inspired cooperative strategies for optimization (NICSO 2010)* (pp. 65-74). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-12538-6\\_6](https://doi.org/10.1007/978-3-642-12538-6_6)

Yang, X.-S. (2012). Flower pollination algorithm for global optimization. In *International conference on unconventional computing and natural computation* (Vol. 7445, pp. 240-249). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-32894-7\\_27](https://doi.org/10.1007/978-3-642-32894-7_27)

Yang, X.-S. (2013). Metaheuristic optimization: Nature-inspired algorithms and applications. In *Artificial Intelligence, Evolutionary Computing and Metaheuristics* (pp. 405-420). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-29694-9\\_16](https://doi.org/10.1007/978-3-642-29694-9_16)

Yang, X.-S. (2014). Swarm intelligence based algorithms: a critical analysis. *Evolutionary intelligence*, 7(1), 17-28. <https://doi.org/10.1007/s12065-013-0102-2>

Yang, X.-S., & Deb, S. (2009, December). Cuckoo Search via Lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)* (pp. 210-214). IEEE. <https://doi.org/10.1109/nabic.2009.5393690>

Yang, X.-S., & He, X. (2013). Firefly algorithm: recent advances and applications. *International Journal of Swarm Intelligence*, 1(1), 36-50. <https://doi.org/10.1504/IJSI.2013.055801>

Yegnanarayana, B. (2009). *Artificial neural networks*. Prentice-Hall India.

Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 856-863).

Zhang, J. G., & Deng, H. W. (2007). Gene selection for classification of microarray data based on the Bayes error. *BMC bioinformatics*, 8(1), 1-9.  
<https://doi.org/10.1186/1471-2105-8-370>

Zhao, C. Y., Zhang, H. X., Zhang, X. Y., Liu, M. C., Hu, Z. D., & Fan, B. T. (2006). Application of support vector machine (SVM) for prediction toxic activity of different data sets. *Toxicology*, 217(2-3), 105-119.  
<https://doi.org/10.1016/j.tox.2005.08.019>

Zhu, Z., Ong, Y. S., & Dash, M. (2007, November). Markov blanket-embedded genetic algorithm for gene selection. *Pattern Recognition*, 40(11), 3236-3248.  
<https://doi.org/10.1016/j.patcog.2007.02.007>



## Appendix A

### Classification Results: SVM vs. CFS-SVM

Table A.1

*Classification Results: SVM vs. CFS-SVM*

Run	SVM				CFS-SVM			
	Colon	Leukemia2	Leukemia3	SRBCT	Colon	Leukemia2	Leukemia3	SRBCT
	Accuracy (%)							
1	73.68	90.91	95.45	96	73.68	95.45	100	100
2	84.21	86.36	95.45	96	78.95	95.45	100	100
3	89.47	95.45	95.45	92	78.95	100	95.45	100
4	52.63	95.45	100	100	89.47	100	95.45	100
5	78.95	100	86.36	96	89.47	100	90.91	100
6	84.21	90.91	100	100	78.95	100	95.45	100
7	89.47	95.45	77.27	100	89.47	100	100	100
8	73.68	81.82	90.91	100	84.21	100	95.45	100
9	84.21	95.45	95.45	92	94.74	100	100	100
10	68.42	100	100	92	84.21	100	100	100
11	78.95	95.45	77.27	100	84.21	95.45	90.91	100
12	84.21	86.36	95.45	100	94.74	100	100	100
13	73.68	100	90.91	100	78.95	100	95.45	100
14	63.16	95.45	72.73	100	84.21	100	100	100
15	89.47	100	95.45	100	94.74	95.45	95.45	100
16	84.21	95.45	72.73	100	84.21	100	100	100
17	78.95	100	90.91	100	84.21	100	95.45	100
18	78.95	100	90.91	96	94.74	100	95.45	100
19	84.21	95.45	90.91	100	84.21	95.45	100	100
20	84.21	100	95.45	100	94.74	100	100	100

## Appendix B

### Classification Results: FA-SVM

Table B.1

*Classification Results: FA-SVM for Colon, Leukemia2, and Leukemia3 Datasets*

Run	Colon		Leukemia2		Leukemia3			
	Subset=1		Subset=1		Subset=1		Subset=2	
	Accuracy (%)							
<b>1</b>	52.63	90.91	81.82	90.91	86.36	86.36	90.91	86.36
<b>2</b>	78.95	95.45	81.82	77.27	81.82	81.82	86.36	90.91
<b>3</b>	57.89	95.45	81.82	90.91	90.91	90.91	95.45	86.36
<b>4</b>	63.16	<b>100</b>	90.91	77.27	95.45	95.45	95.45	<b>100</b>
<b>5</b>	89.47	77.27	86.36	86.36	81.82	95.45	86.36	72.73
<b>6</b>	52.63	90.91	90.91	81.82	90.91	95.45	81.82	90.91
<b>7</b>	84.21	90.91	90.91	72.73	90.91	90.91	81.82	86.36
<b>8</b>	68.42	81.82	90.91	72.73	90.91	95.45	86.36	<b>100</b>
<b>9</b>	63.16	90.91	86.36	95.45	90.91	90.91	90.91	86.36
<b>10</b>	84.21	86.36	90.91	90.91	86.36	90.91	95.45	95.45
<b>11</b>	63.16	<b>100</b>	90.91	86.36	86.36	90.91	72.73	81.82
<b>12</b>	78.95	95.45	86.36	86.36	86.36	86.36	81.82	90.91
<b>13</b>	57.89	59.09	90.91	86.36	77.27	81.82	86.36	90.91
<b>14</b>	68.42	59.09	77.27	72.73	95.45	95.45	77.27	86.36
<b>15</b>	63.16	90.91	86.36	86.36	81.82	86.36	81.82	<b>100</b>
<b>16</b>	73.68	72.73	86.36	95.45	86.36	90.91	90.91	77.27
<b>17</b>	<b>100</b>	95.45	86.36	90.91	72.73	95.45	86.36	77.27
<b>18</b>	63.16	90.91	86.36	95.45	81.82	86.36	95.45	90.91
<b>19</b>	52.63	50	45.45	81.82	90.91	95.45	95.45	90.91
<b>20</b>	52.63	59.09	81.82	95.45	90.91	77.27	86.36	95.45

Table B.2

*Classification Results: FA-SVM for SRBCT Dataset*

Run	SRBCT						
	Subset=1	Subset=2	Subset=3	Subset=4	Subset=5	Subset=6	Subset=7
	Accuracy (%)						
<b>1</b>	60	64	68	76	80	84	92
<b>2</b>	64	64	68	68	80	80	88
<b>3</b>	56	68	92	76	88	96	92
<b>4</b>	64	68	72	76	76	88	96
<b>5</b>	52	80	84	88	84	84	84
<b>6</b>	68	60	88	72	88	88	92
<b>7</b>	68	84	92	76	88	84	80
<b>8</b>	44	72	80	88	88	96	84
<b>9</b>	64	76	88	72	92	88	<b>100</b>
<b>10</b>	60	68	80	84	84	84	92
<b>11</b>	64	80	76	84	84	80	84
<b>12</b>	56	76	76	76	76	88	88
<b>13</b>	60	76	76	76	76	88	84
<b>14</b>	56	76	84	72	84	80	80
<b>15</b>	64	72	84	80	80	80	88
<b>16</b>	68	76	88	76	88	88	92
<b>17</b>	64	76	80	64	80	88	92
<b>18</b>	56	72	76	80	88	84	96
<b>19</b>	68	68	76	84	80	88	92
<b>20</b>	52	80	60	84	84	92	92

## Appendix C

### Classification Results: CFS-FA-SVM

Table C.1

*Classification Results: CFS-FA-SVM*

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Subset=1	Subset=2	Subset=1	Subset=2	Subset=1	Subset=2	Subset=3	Subset=4
	Accuracy (%)							
<b>1</b>	89.47	84.21	86.36	<b>100</b>	48	80	76	<b>100</b>
<b>2</b>	84.21	94.74	95.45	86.36	56	84	84	<b>100</b>
<b>3</b>	73.68	94.74	95.45	81.82	68	92	88	92
<b>4</b>	68.42	94.74	90.91	90.91	64	84	84	<b>100</b>
<b>5</b>	73.68	89.47	81.82	90.91	64	84	80	96
<b>6</b>	73.68	78.94	86.36	95.45	68	88	76	<b>100</b>
<b>7</b>	73.68	94.74	90.91	86.36	56	88	92	88
<b>8</b>	78.95	89.47	<b>100</b>	90.91	60	76	80	88
<b>9</b>	89.47	94.74	95.45	86.36	60	80	92	96
<b>10</b>	89.47	84.21	95.45	90.91	64	84	92	92
<b>11</b>	63.16	89.47	90.91	86.36	72	88	80	92
<b>12</b>	84.21	<b>100</b>	95.45	81.82	68	72	88	96
<b>13</b>	84.21	89.47	86.36	81.82	64	72	96	96
<b>14</b>	57.89	89.47	81.82	90.91	72	80	88	96
<b>15</b>	78.95	94.74	81.82	95.45	56	84	88	92
<b>16</b>	94.74	84.21	86.36	77.27	52	68	92	96
<b>17</b>	84.21	84.21	77.27	90.91	52	76	64	96
<b>18</b>	73.68	84.21	81.82	90.91	60	84	92	<b>100</b>
<b>19</b>	73.68	94.74	81.82	77.27	52	76	92	92
<b>20</b>	73.68	84.21	90.91	86.36	60	76	80	96

## Appendix D

### Classification Results: CFS-MCFA-SVM

Table D.1

*Classification Results: CFS-MCFA-SVM with FF-SVM Parameters*

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Accuracy (%)	Subset Size						
<b>1</b>	100	4	100	2	100	6	<b>100</b>	<b>6</b>
<b>2</b>	89.4737	3	<b>100</b>	<b>1</b>	100	8	100	10
<b>3</b>	94.7368	3	100	2	100	7	100	9
<b>4</b>	100	9	100	5	100	10	100	8
<b>5</b>	89.4737	8	100	4	100	5	100	7
<b>6</b>	100	8	100	6	100	7	100	8
<b>7</b>	100	10	100	2	100	6	100	7
<b>8</b>	<b>100</b>	<b>3</b>	100	2	100	11	100	12
<b>9</b>	94.7368	4	100	4	<b>100</b>	<b>4</b>	100	9
<b>10</b>	89.4737	7	100	1	100	6	100	8
<b>11</b>	100	3	100	2	100	10	100	7
<b>12</b>	94.7368	8	100	5	100	9	100	8
<b>13</b>	100	5	100	4	100	16	100	7
<b>14</b>	100	9	100	6	100	8	100	9
<b>15</b>	89.4737	8	100	2	100	7	100	8
<b>16</b>	89.4737	4	100	2	100	10	100	7
<b>17</b>	94.7368	3	100	4	100	5	100	8
<b>18</b>	100	7	100	1	100	7	100	7
<b>19</b>	94.7368	7	100	4	100	6	100	12
<b>20</b>	84.2105	6	100	2	100	11	100	9

Table D.2

Classification Results: CFS-MCFA-SVM with Co-ABC Parameters

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Accuracy (%)	Subset Size						
1	100	4	100	2	100	5	100	4
2	100	9	100	1	100	5	100	10
3	100	1	100	1	100	8	100	6
4	100	3	100	2	100	9	100	5
5	100	3	100	1	100	2	100	11
6	100	9	100	3	100	6	100	8
7	100	9	100	1	100	9	100	6
8	100	5	100	2	100	2	100	10
9	94.7368	3	100	1	100	6	100	6
10	94.7368	2	100	2	100	9	100	7
11	100	4	100	3	100	2	100	10
12	100	8	100	1	100	6	100	6
13	100	4	100	3	100	9	100	7
14	100	7	100	1	100	2	100	9
15	100	3	100	2	100	6	100	9
16	100	2	100	1	100	4	100	10
17	100	9	100	2	100	5	100	10
18	100	5	100	3	100	6	100	6
19	94.7368	3	100	3	100	4	100	10
20	94.7368	2	100	2	100	5	100	6
21	100	9	100	3	100	5	100	7
22	100	5	100	1	100	9	100	9
23	94.7368	3	100	3	100	2	100	9
24	94.7368	2	100	1	100	6	100	10
25	100	4	100	2	100	4	100	6
26	100	5	100	1	100	9	100	7
27	100	5	100	2	100	2	100	10
28	100	4	100	3	100	6	100	6
29	89.4737	5	100	3	100	4	100	7
30	100	9	100	1	100	5	100	9

Table D.3

*Classification Results: CFS-MCFA-SVM with FFF-SVM Parameters*

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Accuracy (%)	Subset Size						
1	100	4	100	2	100	6	100	6
2	89.4737	3	<b>100</b>	<b>1</b>	100	8	100	10
3	94.7368	3	100	2	100	7	100	9
4	100	9	100	5	100	10	100	8
5	89.4737	8	100	4	100	5	100	7
6	100	8	100	6	100	7	100	8
7	100	10	100	2	100	6	100	7
8	<b>100</b>	<b>3</b>	100	2	100	11	100	12
9	94.7368	4	100	4	<b>100</b>	<b>4</b>	100	9
10	89.4737	7	100	1	100	6	100	8
11	100	3	100	2	100	10	100	7
12	94.7368	8	100	5	100	9	100	8
13	100	5	100	4	100	16	100	7
14	100	9	100	6	100	8	100	9
15	89.4737	8	100	2	100	7	100	8
16	89.4737	4	100	2	100	10	100	7
17	94.7368	3	100	4	100	5	100	8
18	100	7	100	1	100	7	100	7
19	94.7368	7	100	4	100	6	100	12
20	84.2105	6	100	2	100	11	100	9
21	100	8	100	2	100	4	100	8
22	89.4737	5	100	3	100	6	100	7
23	94.7368	5	100	2	100	6	<b>100</b>	<b>5</b>
24	100	6	100	2	100	5	100	8
25	84.2105	3	100	4	100	4	100	6

Table D.4

Classification Results: CFS-MCFA-SVM with rMRMR-MBA Parameters

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Accuracy (%)	Subset Size						
1	100	4	100	3	100	4	100	8
2	100	10	<b>100</b>	<b>1</b>	100	4	100	7
3	<b>100</b>	<b>2</b>	100	2	100	7	100	6
4	94.736	6	100	2	100	7	100	6
5	100	6	100	3	100	5	<b>100</b>	<b>3</b>
6	94.736	11	100	2	100	8	100	8
7	100	7	100	1	100	9	100	6
8	94.736	9	100	2	100	7	100	7
9	94.736	8	100	2	100	7	100	8
10	100	6	100	3	100	5	100	6
11	100	10	100	2	100	8	100	8
12	100	2	100	2	100	9	100	7
13	94.736	6	100	1	100	5	100	8
14	100	6	100	2	100	6	100	6
15	100	4	100	2	100	5	100	7
16	100	9	100	3	100	3	100	7
17	100	3	100	2	100	7	100	8
18	100	7	100	2	100	6	100	7
19	100	8	100	2	100	3	100	6
20	100	10	100	3	100	7	100	7
21	100	2	100	2	100	4	100	7
22	94.736	6	100	1	100	5	100	6
23	100	6	100	2	100	3	100	7
24	100	10	100	2	100	3	100	7
25	100	2	100	3	<b>100</b>	<b>2</b>	100	8
26	94.736	6	100	2	100	4	100	7
27	100	8	100	2	100	4	100	9
28	94.736	2	100	2	100	6	100	7
29	94.736	5	100	3	100	5	100	7
30	100	11	100	2	100	5	100	7

Table D.5

*Classification Results: CFS-MCFA-SVM with FFF-SVM Parameters using LOOCV*

Run	Colon		Leukemia2		Leukemia3		SRBCT	
	Accuracy (%)	Subset Size						
1	100	22	100	3	100	4	100	10
2	100	12	100	3	100	4	100	8
3	100	18	100	3	100	6	100	7
4	100	24	100	3	100	11	100	7
5	100	15	100	2	100	5	100	10
6	100	14	100	5	100	6	100	8
7	100	26	100	4	100	5	100	8
8	100	26	100	3	100	5	100	7
9	100	7	100	2	100	4	100	8
10	100	26	100	6	100	5	100	15
11	100	7	100	3	100	4	100	7
12	100	23	100	3	100	5	100	10
13	100	26	100	3	100	7	100	8
14	100	7	100	3	100	6	100	8
15	100	14	100	3	100	6	100	7
16	100	10	100	2	100	4	100	8
17	100	4	100	5	100	4	100	15
18	100	13	100	2	100	5	100	9
19	100	7	100	4	100	7	100	12
20	100	14	100	2	100	6	100	9
21	100	26	100	2	100	4	100	9
22	100	7	100	3	100	5	100	9
23	100	26	100	4	100	7	100	11
24	100	7	100	2	100	6	100	9
25	100	23	100	4	100	4	100	8

## Appendix E

### Classification Results: Co-ABC-SVM

Table E.1

*Classification Results: Co-ABC-SVM*

Run	Colon		Leukemia2		Leukemia3		SRBCT		
	Subset=1		Subset=1		Subset=1		Subset=2	Subset=1	Subset=2
	Accuracy (%)	Subset=3							
1	78.95	95.45	86.36	77.27	56	84	92		
2	100	95.45	72.73	86.36	56	68	96		
3	68.42	90.91	81.82	90.91	60	76	92		
4	89.47	90.91	77.27	100	60	80	96		
5	68.42	90.91	77.27	86.36	60	80	84		
6	78.95	95.45	86.36	81.82	56	80	100		
7	89.47	86.36	86.36	90.91	44	80	84		
8	94.74	95.45	77.27	86.36	68	84	92		
9	84.21	90.91	95.45	86.36	56	84	92		
10	73.68	90.91	86.36	90.91	60	84	92		
11	84.21	86.36	81.82	86.36	60	84	80		
12	78.95	81.82	90.91	90.91	56	80	92		
13	89.47	90.91	86.36	86.36	60	88	96		
14	84.21	90.91	77.27	90.91	52	88	84		
15	73.68	100	72.73	81.82	44	84	96		
16	94.74	86.36	81.82	86.36	68	80	92		
17	78.95	100	81.82	77.27	48	88	100		
18	73.68	90.91	86.36	95.45	72	88	88		
19	78.95	95.45	90.91	86.36	60	88	72		
20	94.74	90.91	81.82	95.45	60	84	92		

Table E.1 continued

<b>21</b>	68.42	81.82	81.82	<b>100</b>	60	88	92
<b>22</b>	63.16	90.91	86.36	<b>100</b>	72	84	84
<b>23</b>	73.68	90.91	81.82	86.36	52	80	96
<b>24</b>	94.74	90.91	86.36	86.36	60	84	92
<b>25</b>	78.95	86.36	81.82	90.91	60	80	96
<b>26</b>	73.68	90.91	77.27	81.82	52	80	84
<b>27</b>	78.95	86.36	95.45	86.36	48	80	<b>100</b>
<b>28</b>	94.74	90.91	72.73	77.27	60	84	84
<b>29</b>	68.42	<b>100</b>	77.27	95.45	64	80	92
<b>30</b>	63.16	90.91	81.82	86.36	64	76	92

## Appendix F

### Publications during the Study

#### Journals

1. Fajila, F., & Yusof, Y. (2021, April 1). Incremental Search for Informative Gene Selection in Cancer Classification. *Annals of Emerging Technologies in Computing (AETiC)*, 5(2), 15-21.  
<https://doi.org/10.33166/AETiC.2021.02.002>

Indexed by Scopus

2. Fajila, M. N. F., & Yusof, Y. (in press). Hybrid gene selection with mutable firefly algorithm for feature selection in cancer classification. *International Journal of Intelligent Engineering and Systems (IJIES)*

Indexed by Scopus