

Northumbria Research Link

Citation: Dreder, Abdouladeem, Tahir, Muhammad, Seker, Huseyin and Anwar, Naveed (2016) Discovering differences in gender-related skeletal muscle aging through the majority voting-based identification of differently expressed genes. International Journal on Bioinformatics & Biosciences, 6 (2). pp. 1-14. ISSN 1839-9614

Published by: WSP

URL: <http://wireilla.com/papers/ijbb/V6N2/6216ijbb01.pdf>
<<http://wireilla.com/papers/ijbb/V6N2/6216ijbb01.pdf>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/27774/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

www.northumbria.ac.uk/nrl



DISCOVERING DIFFERENCES IN GENDER-RELATED SKELETAL MUSCLE AGING THROUGH THE MAJORITY VOTING-BASED IDENTIFICATION OF DIFFERENTLY EXPRESSED GENES

Abdouladeem Dreder^{1,2}, Muhammad Tahir¹, Huseyin Seker¹ and Muhammad Anwar¹

¹Bio-Health Informatics Research Group, Faculty of Engineering and Environment, The University of Northumbria at Newcastle, Newcastle-upon-Tyne, UK

²Biotechnology Research Centre, P. O. Box 30313, Tripoli, Libya

ABSTRACT

Understanding gene function (GF) is still a significant challenge in system biology. Previously, several machine learning and computational techniques have been used to understand GF. However, these previous attempts have not produced a comprehensive interpretation of the relationship between genes and differences in both age and gender. Although there are several thousand of genes, very few differentially expressed genes play an active role in understanding the age and gender differences. The core aim of this study is to uncover new biomarkers that can contribute towards distinguishing between male and female according to the gene expression levels of skeletal muscle (SM) tissues. In our proposed multi-filter system (MFS), genes are first sorted using three different ranking techniques (t-test, Wilcoxon and Receiver Operating Characteristic (ROC)). Later, important genes are acquired using majority voting based on the principle that combining multiple models can improve the generalization of the system. Experiments were conducted on Micro Array gene expression dataset and results have indicated a significant increase in classification accuracy when compared with existing system.

INDEX TERMS

Multi-Filter System, Filter Techniques, Micro Array Gene Expression, Skeletal muscle

I. INTRODUCTION

Sexual dimorphism of skeletal muscle can occur due to age [1] and many of these age-related changes in skeletal muscle appear to be influenced by gender [2], [17], [18]. For example, the muscle mass of men is larger than that of women, especially for type II fibers, while the type I muscle fibers proportion of oxidative is higher in women [3]. Welle et al. reported that the muscle mass of men is larger than that of women [1], [11], [12], due to the higher level of testosterone and the anabolic effect of testosterone is well known. However, previous studies have failed to identify which genes are responsible for anabolic effects. The molecular biases related to gender difference are still fuzzy [1]; 50% of the cell mass of the human body is muscle, so skeletal muscle is considered an important issue. There are several changes in skeletal muscle related to age that seem to be influenced by gender [4]. These changes in gene expression could be responsible for the decline in muscle function [5]. In relation to sex, despite the fact that there are higher number of genes in expression related to gender difference, very few genes can help to interpret the gender difference issue [3]. For the profiles of men and women, there are few comparisons of broad gene expression that have been carried out [5].

Janssen et al [13] reported that the reduction of skeletal muscle (SM) mass related to age starts in the third decade. This decrease starts to appear in the lower body SM. To find differences between men and women, they used t-test, Pearson correlation and multiple regression to determine the

relationship between age and skeletal muscle. Liu et al [2] used basic statistical analysis to make a comparison between males and females in each set of age using gene expression profiles from skeletal muscle tissue. They identified important sex and age related gene functional groups using intensity-based Bayesian moderated t-test and logistic regression. This was the first study that offers global proof for the occurrence of extensive sex changes in the aging process of human skeletal muscle. Although the study showed interesting results, but they had used genes belonging to X and Y chromosomes, which can easily discriminate genders. Experiments were conducted using 3 groups namely older women versus old men, young women versus older women, and young men versus older men. But the main problem in their study is that important genes are identified using whole training data. This can lead to poor generalization because one of the fundamental goal of machine learning is to generalize beyond the samples in the training data.

The main goal of this study is to extend the work reported by Liu et al [2] by identifying important genes with good generalization ability. So in this work we prefer to applied work reported by Liu et al. [2], propose a method and make a comparison with them in order to show the effectiveness of our approach using two datasets which will be explained in proposed method and material section.

In our proposed approach which is basically inspired from ensemble of feature ranking methods for data intensive application [16], genes are first sorted using three different ranking techniques (t-test, Wilcoxon and ROC). Later, important genes are acquired using majority voting based on the principle that combining multiple models can improve the generalization of the system. The scope of this paper is the selection of the most reliable genes and the evaluation of classification power of selected genes. Experiments were conducted on Micro Array gene expression dataset and results have indicated a significant increase in classification accuracy when compared with the genes obtained by the system in [2]. In this study we applied our proposed technique on two data sets and our system is able to identify differentially-expressed genes for the following three case studies in relation to age and gender differences

- Young Women versus Old Women
- Young Men versus Old Men
- Old Men versus Old Women

This paper is organized as follows. Section II describes material and the proposed method followed by results and discussion in Section III. Section IV concludes the paper.

II. MATERIAL AND PROPOSED METHOD

A. Micro array gene expression data set

In this study, two datasets contain a microarray dataset of gene expression of skeletal muscle atissue. Datasetsare publicly available in the Gene Expression Omnibus (GEO) database [2], [19]. A total of 58 individuals were involved in this investigation:

22 healthy males and females of various ages, in which 7 males and 7 females were young (20-29 years old), and 4 males and 4 females were old (61-81 years old), were included in studyA. The whole Ribonucleic Acid (RNA) was extracted and gene expression profiling was implemented utilizing Affymetrix human genome U133 Plus 2chip. As in [2], this data set is divided into three cases, the first case involved 11 females (7 young and 4 old), second case consist of 11 males (7 young and 4 old) and the last case contain of 8 samples (4 old men and 4 old women).In study B,genes subset selection using Feature ranking techniques Bioinformatics data have extremely high dimensionality, and around 55,000 genes with only 36 samples, 15 young (7 men, 8 women) and 21 old (10 men and 11 women) were included in study B.

B. Genes subset selection using Feature ranking techniques

The datasets used in this study have extremely high dimensionality. The first dataset consists of around 55,000 genes with only 22 samples. While the second dataset [19] involves 55,000 genes with 36 samples. This is considered a significant challenge to machine learning methods, this means that there are a large number of features than samples. To address this problem, it is important to select small relevant features subset to reduce processing time and avoid over fitting problem [6]. One of the possible solution is feature selection using feature ranking methods. In this study, three different filter methods are investigated. These methods are summarized as follows

- **t-test:** a statistical hypothesis where the statistic follows a Student's t distribution [9]. It is usually used to evaluate if the averages of two classes are not statistically similar by computing the variability and difference between two classes.
- **Entropy:** is normally used for high dimensional data to select the suitable number of features using the principle of Entropy. In this method the distance between the probability density functions is measured by divergence, which means that the features with higher divergence are considered more suitable for discriminating classes [20].
- **Receiver Operating Characteristic (ROC):** offers an active method to characterize the classifier sensitivity versus specificity. It is drawn between sensitivity and 1-specificity, for different values of the threshold, and based on the area under the ROC curve, ranking of the features is performed [20,21].

C. Classification

Selected subset of genes are tested for its generalization power using supervised classification. k-nearest neighbor (KNN) classifier (k=1,3) is used to evaluate the system performance. The leave-one-out cross validation (LOOCV) technique is used for evaluation.

D. Proposed System

Figure I shows the framework of the proposed system which is inspired from the fact that combining multiple models can improve the generalization of the system. We first divided the data set using leave-one-out-cross validation into T folds. In other words, there are 20 folds for 20 samples where each fold consists of 19 samples for training and one sample for testing. For each fold multi filter system (MFS) is applied, which includes three different rank feature filters T-test, ROC and Entropy. Each filter ranking technique is responsible to sort genes according to criteria specified in the filter ranking methods. From these sorted genes, N unique subset of genes are obtained based on majority voting. Which is shown in Table I. Let's assume that there are total 10 genes and the objective is to select top 5 genes. Genes 9 and 10 are selected by all feature ranking techniques so these are most important genes. Genes 1, 4, 5 are selected twice and thus are also considered as important genes by the system. It should be noted that due to majority voting genes 2, 3 and 6 are not selected by the system. Later, KNN is applied on the new subset of genes in order to check the predictive performance.

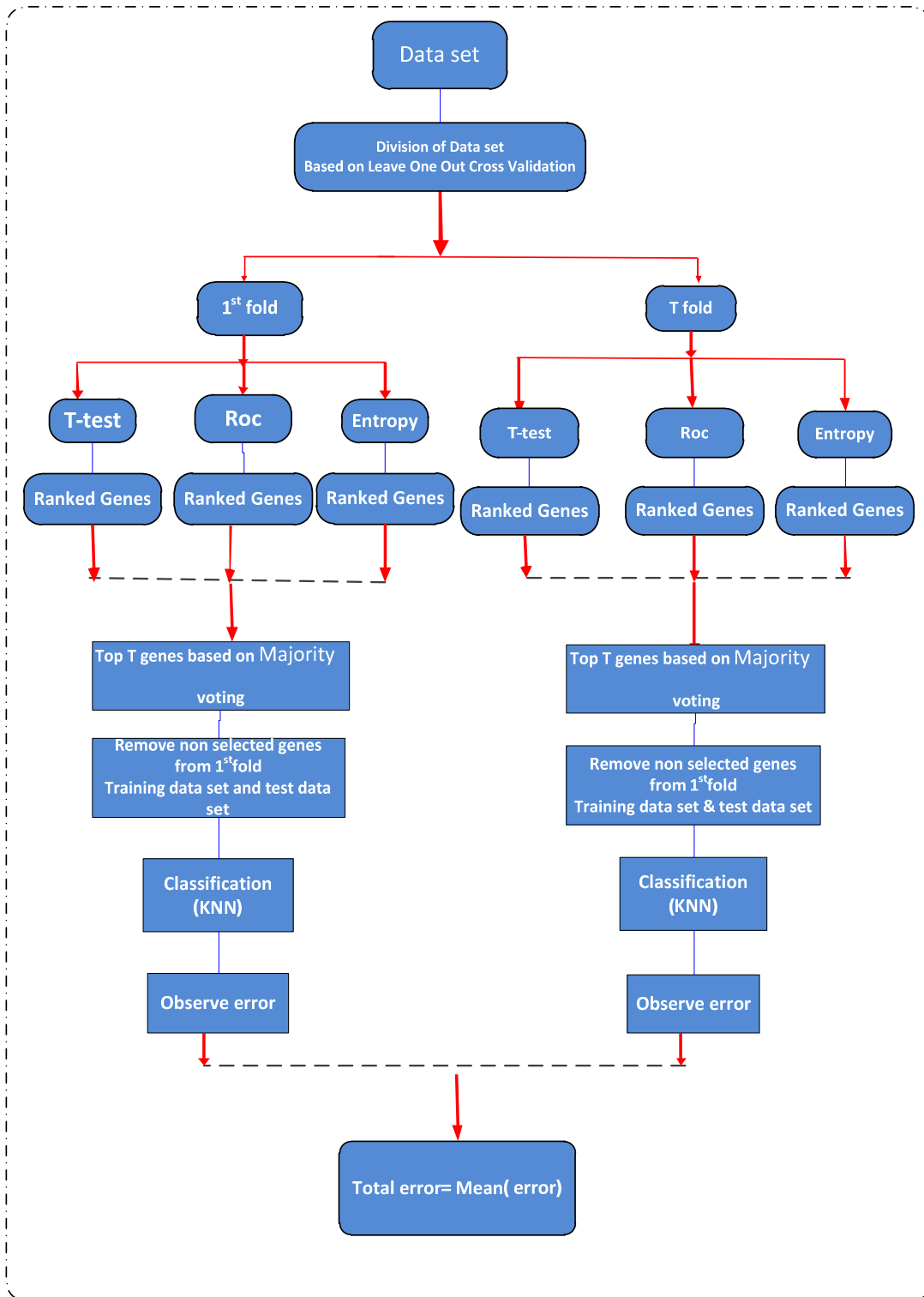


Fig. I. Proposed Multi-Filter System (MFS).

TABLE I : Majority Voting to Select Important Genes.

	Top Ranked Genes
t-test	1, 4, 6, 9, 10
ROC	1, 2, 5, 9, 10
Entropy	3, 4, 5, 9, 10
Majority Voting	1, 4, 5, 9, 10

k-nearest neighbor: The main objective of k-nearest neighbor (k-NN) classifier is to discover set of k objects in the training set that are similar to the objects in the testgroup [14]

$$d(x_i - x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

where a is the feature vector of xth sample.

III. RESULTS AND DISCUSSION

In this section, we will evaluate the performance of the multi-filter system (MFS). The proposed system is also compared with the system presented in [2] and [19], in which genes are identified for three categories (male young versus male old, female young versus female old and male old versus female old) from total of 54623 genes. In order to have a fair comparison, the same number of genes are selected from MFS and compared with the genes identified in [2] and [19]. The evaluation metrics used in this study are: Classification accuracy, Sensitivity and Specificity.

The main purpose of confusion matrix is to display the percentage of correctly classified true positive (TP) and true negative (TN) objects and false classified false positive (FP) and false negative (FN) objects as shown in TableXIII:

TABLE II: Confusion matrix for two-class classification problem.

	A	B
Real class	True Positive TP	False Negative FN
	False Positive FP	True Negative TN

- True positive rate (TP)=TP/(TP+FN)
- False negative rate (FN)=FN/(TP+FN)=1- Sensitivity
- False positive rate (FP) =FP/(FP+TN)= 1- Specificity.

Study A

In the present study the dataset contains microarray dataset which is publicly available at the Gene Expression Omnibus (GEO) dataset <http://www.ncbi.nlm.nih.gov/>, for use by the scientific community: [GSE38718](#).

Subjects, the subjects were 22 healthy male and female in various ages, young 11 males and 11 females (20-29 years old) and 8 old subjects (4 males and 4 females) all of the (61-81 years old). Total RNA was extracted and gene expression profiling was performed using the Affymetrix Human Genome U133 plus 2 chip.

A. Case Study 1: Young Men versus Old Men

This case study consists of 11 male samples (7 young and 4 old). Table II shows the performance of MFS when compared with the genes identified by Liu et al [2]. It is observed that the best performance is obtained using 3NN classifier which is 90.9% while genes obtained by [2] only able to achieve 81.8%. This improvement is mainly due to high specificity. Further analysis has revealed that out of 75 genes, only 9 genes are common in both systems. Some new genes are identified, that can play an important role in age differences of young and old males. Some of the new genes are shown in Table III along with 9 genes that are selected by both systems.

These new genes can be very useful for biologist in order to identify the differences between young and old males. Figure II, III, IV, V, VI, VII are shown the performance of the system by varying the number of genes. Compared with [2]. It is observed that the best performance is obtained by using 10 or 20 genes and afterwards, there is a 10% drop in performance. This may be due to selection of some genes that can degrade the performance of the system. Future work aims to investigate wrapper techniques to identify these genes.

TABLE III : Young Men Versus Old Men.

classifier	Correct Rate		Sensitivity		Specificity	
	M FS	2	M FS	2	M FS	2
KNN(K=1)	0. 818	0. 636	0. 714	0.5 714	1 .0	0. 75
KNN(K=3)	0. 909	0. 818	0. 857	0.8 57	1 .0	0. 75

TABLE IV : Young men versus old men. New genes selected by the proposed system. Common genes selected by proposed system and system by [2].

New genes selected by MFS Common Genes	New genes selected by MFS Common Genes
<ul style="list-style-type: none"> • Caveolin 3 (CAV3) • Eukaryotic translation elongation (EEF1B2) • FBR-MuSV ubiquitously expressed (FAU) • RNA binding motif protein 15 (RBM15) • Ribosomal protein L4 (RPL4) • Cytochrome c-1 (CYC1) • Mitochondrial ribosomal protein S30 (MRPS30) • Pyruvate dehydrogenase kinase, isozyme 2 (PDK2) • Phosphoglycerate mutase 2 muscle (PGAM2) 	<ul style="list-style-type: none"> • Toll-like receptor 4 (TLR4) • UDP-GlcNAc:betaGal (B3GNT6) • TGF-beta activated kinase 1 (TAB3) • Myozenin 3 (MYOZ3) • Olfactory Receptor (OR5P3) • Thioesterase superfamily member 4 (THEM4) • RAN binding protein 3-like (RANBP3L) • Fc receptor-like 3 (FCRL3) • Rhomboid, veinlet-like 3 Drosophila (RHBDL3)

B. Case Study 2: Old Men versus Old Women

Another objective of this investigation was to examine basal level gene expression among Old Men and Old Women. In This case study consists of 8 adults (4 old men versus 4 old women). Table IV shows the performance of MFS when compared with the genes identified by Liu et al. [2]. It is observed that genes selected using MFS have classification accuracy of 100% using both 1NN and 3NN with high Sensitivity and Specificity.

TABLE V Old Men Versus Old Women.

classifier	Correct Rate		Sensitivity		Specificity	
	M	2	M	2	M	2
	FS		FS		FS	
KNN(K=1)	1	0	1	0	1	1
	.0	.75	.0	.5	.0	.0
KNN(K=3)	1	0	1	0	1	1
	.0	.375	.0	.5	.0	.0

C. Case Study 3: Young Female versus Old Female

This case study consists of 11 female samples (7 young and 4 old). Table V shows the performance of MFS when compared with the genes identified by Liu et al [2]. Again, the best performance is obtained using 1NN classifier which is 91%. While genes identified by [2] are only able to achieve 72.2% accuracy which indicates the important improved generalization ability of the proposed system. We argue that improvement in performance is mainly due to high Specificity as Sensitivity which is same in the both systems.

TABLE VI. Young Female Versus Old Female.

classifier	Correct Rate		Sensitivity		Specificity	
	M FS	2	MF S	2	M FS	2
KNN(K=1)	0. 91	0. 454	0.8 571	0. 571	1. 0	. 25
KNN(K=3)	0. 722	0. 722	0.8 57	0. 857	0. 5	0 .5

Study B

This dataset consist of 54,623 genes expression dataset, it is publicly available at the Gene Expression Omnibus (GEO) dataset <http://www.ncbi.nlm.nih.gov/>. The subjects were 36 healthy young males and females in various ages, young (19 female (8 young, 11 old) and (17male (7young, 10 old)) Total RNA was extracted and gene expression profiling was performed using the Affymetrix Human Genome U133 plus 2 chip

Case Study 1: Young male versus Old male

This case of study involves 17 male samples (7 young and 10 old). Table VII shows the comparison of performance of MFS (with second dataset) with the genes identified by Raue et al [19]. According to the information in Table1, the best performance is obtained using 1NN and 3NN classifier which is 88% when genes were obtained by Raue et al [19]are only able to achieve 0.82%. There are 39 genes are common in both systems. Some new genes are identified, that can play an important role in age differences of young and old males. TableVII lists the name of some new genes identified by the proposed system along with some common genes selected by both systems

TABLE VII: Young Male VERSUS Old Male.

classifier	Correct Rate		Sensitivity		Specificity	
	MFS	3	MFS	3	M FS	3
KNN(K=1))	0.88	0.82	0.85	0.71	0. 90	0 .90
KNN(K=3))	0.88	0.82	0.85	0.71	0. 90	0 .90

TABLE VIII: List Of Common And Different Genes Between Raue Et Al [19] And Mfs

¹ Different genes	² Common genes
³ BC016339	⁴ ATP1B4
⁵ DNAJC5B	⁶ CDKN1A
⁷ VPS18	⁸ CLIC5
⁹ ATRNL1	¹⁰ DDB2
¹¹ TRIM40	¹² DMRT2
¹³ ZNF791	¹⁴ DMRT2
¹⁵ ATRNL1	¹⁶ EPB41L3
¹⁷ DNAJC5B	¹⁸ EPB41L3
¹⁹ LHFPL3-AS1	²⁰ EPB41L3
²¹ KRTAP19-1	²² FAM171A1

Case Study 2: Young Female versus Old Female

This case of study involves 19 male samples (8 young and 11 old). Table IX shows the comparison of performance of MFS (with second dataset) with the genes identified by Raue et al [19], the best performance is obtained using 3NN classifier which is 100% while genes obtained by [19] only able to achieve 81.8%. This improvement is mainly due to high specificity and Sensitivity (approximately 100%). Further analysis has shown that out of 102 genes, only 8 genes are common in both systems. Some new genes are identified, that can play an important role in age differences of young and old males. Table X lists the name of some new genes identified by the proposed system along with some common genes selected by both systems

TABLE IX Young Female versus Old Female.

classifier	Correct Rate		Sensitivity		Specificity	
	M FS	3	M FS	3	M FS	3
KNN(K=1)	1	0.89	1	0.87	1	0.90
KNN(K=3)	1	0.84	1	0.62	1	1

TABLE X List Of Common And Different Genes Between Raue Et Al [19] And Mfs

¹ Different genes	² Common genes
³ CIQB	⁴ LOC401220
⁵ DDB2	⁶ LUM
⁷ FEZ2	⁸ MYLK4
⁹ KLF5	¹⁰ MAP2K1
¹¹ PLAG1	¹² BC016339
¹³ SKAP2	¹⁴ TXNRD2
¹⁵ ZNF385B	¹⁶ HPGD
¹⁷ PDK4	¹⁸ RBM25

Case Study 3: Old Men versus Old Women

This case study involves 21 male samples (10 men and 11 women). Table XI shows the comparison of performance of MFS (with second dataset) with the genes identified by Raue et al [19]. According to the Table 1, the best performance is gained using 3NN classifier which is approximately 95% while genes obtained by [19] were only able to achieve 76%. Only 9 genes are common in both systems. Some new genes are identified, that can play an important role in age differences of young male and old male. Some new genes are shown in Table X II lists the name of some new genes identified by the proposed system along with some common genes selected by both systems.

TABLE XI : old male VERSUS old female.

classifier	Correct Rate		Sensitivity		Specificity	
	MFS	3	MFS	3	MF S	3
KNN(K=1)	0.95	0.61	1.00	0.63	0.9	0.60
KNN(K=3)	0.80	0.76	0.71	0.81	0.9	0.70

TABLE X II list of common and different genes between Raue et al [19] and mfs.

¹⁹ Different genes	²⁰ Common genes
²¹ PCDH15	²² CCNG2
²³ BC022384	²⁴ COL4A6
²⁵ DDX3Y	²⁶ FLRT2
²⁷ RPS4Y1	²⁸ FOLR2
²⁹ EIF1AY	³⁰ MS4A4A
³¹ DDX3Y	³² NMNAT1
³³ USP9Y	³⁴ SLPI
³⁵ KDM5D	³⁶ TPPP3
³⁷ AW052159	³⁸ PDK4

IV. CONCLUSION

In this study, multi-filter system (MFS) is proposed to identify important genes for Males and Females using skeletal muscle. Genes are first sorted using three different ranking techniques (t-test, Wilcoxon and ROC). The Proposed system is evaluated on publicly available microarray datasets of gene expression of skeletal muscle tissue.

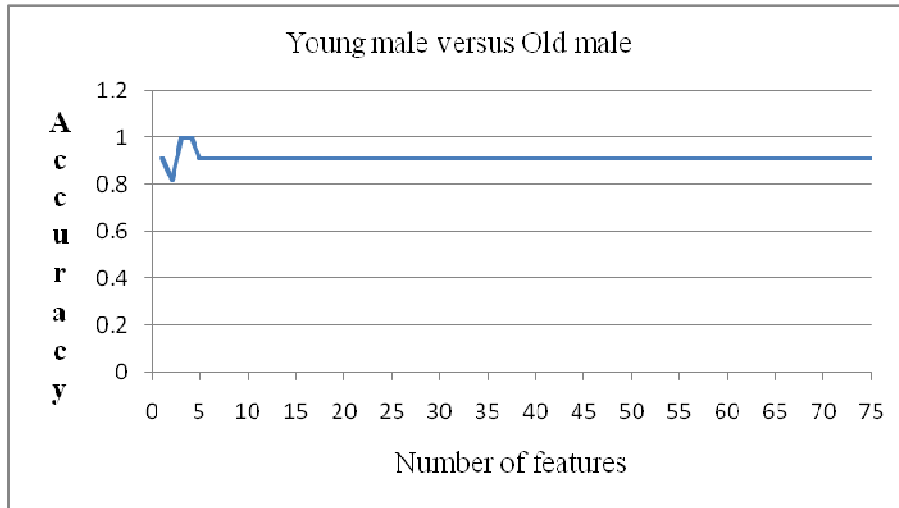


Fig. II. Graph showing performance of the system using varying number of genes compared with[2]

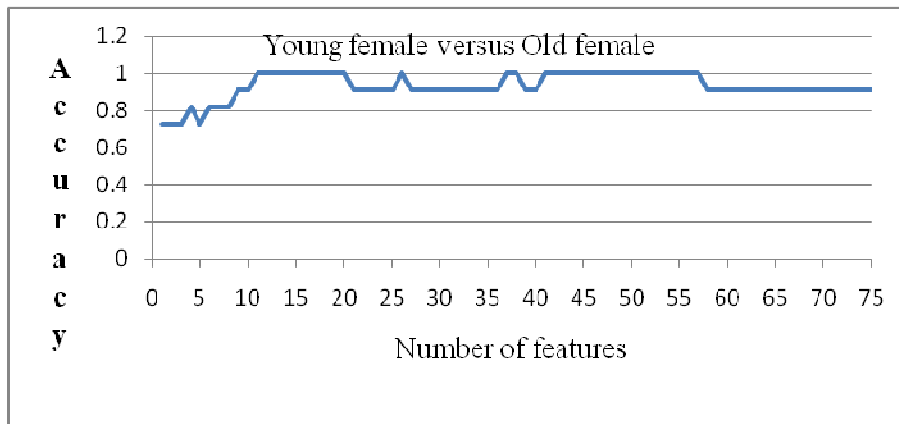


Fig. III. Graph showing performance of the system using varying number of genes compared with[2]

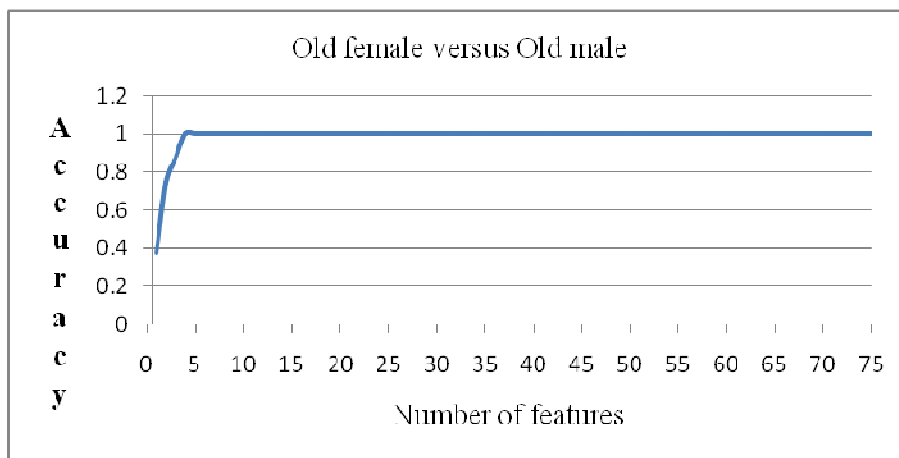


Fig. IV. Graph showing performance of the system using varying number of genes compared with[2]

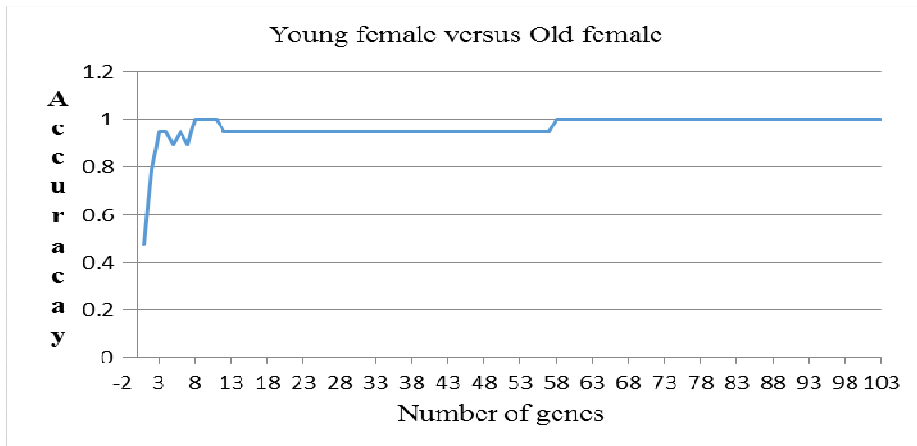


Fig. V. Graph showing performance of the system using varying number of genes compared with [19]

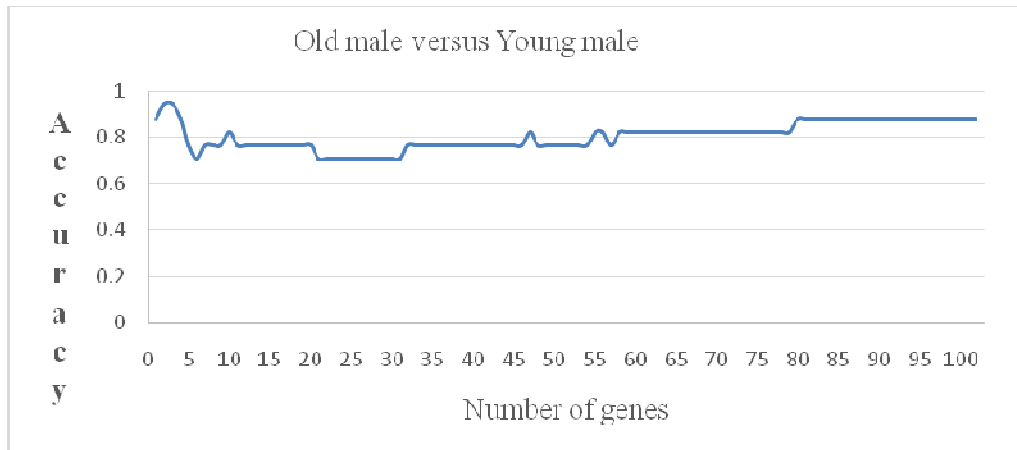


Fig. VI. Graph showing performance of the system using varying number of genes compared with [19]

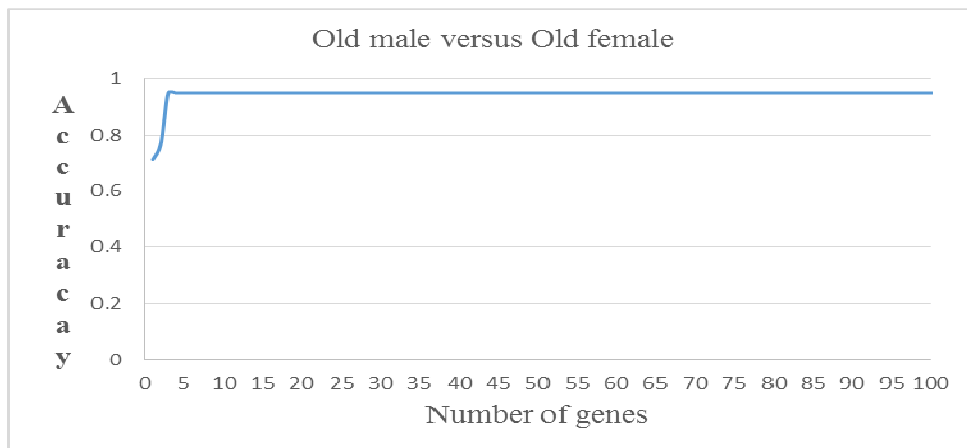


Fig. VII Graph showing performance of the system using varying number of genes compared with [19]

Later, important genes are acquired using majority voting based on the principle that combining multiple models can improve the generalization of the system. The results have indicated that the classification performance achieved by the proposed system yields the best classification performance when compared with similar number of genes identified in previous study [2][19]. Future work aims to improve the performance by identifying more important genes through Wrapper Feature Ranking techniques rather than filter based feature ranking techniques.

REFERENCES

- [1] S. Welle, R. Tawil, and C. A. Thornton, "Sex-related differences in gene expression in human skeletal muscle", *PLoS One*, vol. 3, no. 1, pp. e1385-e1385, 2008
- [2] D. Liu, M. A. Sartor, G. A. Nader, E. E. Pistilli, L. Tanton, C. Lilly, et al., "Microarray analysis reveals novel features of the muscle aging process in men and women", *Biological Sciences*, vol. 68, no.9, pp. 1035-1044, 2013
- [3] D. D. Liu, M. A. Sartor, G. A. Nader, L. Gutmann, M. K. Treutelaar, E. E. Pistilli, H. B. IglayReger, C. F. Burant, E. P. Hoffman, and P. M. Gordon, "Skeletal muscle gene expression in response to resistance exercise: sex specific regulation", *BMC genomics*, vol. 11, no. 1, pp. 659, 2010.
- [4] G. Sifakis, I. Valavanis, O. Papadodima, and A. A. Chatziioannou, "Identifying Gender Independent Biomarkers Responsible for human Muscle Aging Using Microarray Data", *Bioinformatics and Bioengineering (BIBE)*, no. pp. 1-5, 2013
- [5] S. M. Roth, R. E. Ferrell, D. G. Peters, E. J. Metter, B. F. Hurley, and M. A. Rogers, "Influence of age, sex, and strength training on human muscle gene expression determined by microarray", *Physiological genomics*, vol. 10, pp. 181-190, 2002.
- [6] Y. Saeyns, I. a. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics", *bioinformatics*, vol. 23, no. pp. 2507-2517, 2007.
- [7] Y. Su, T. M. Murali, V. Pavlovic, M. Schaffer, and S. Kasif, "RankGene: identification of diagnostic genes based on expression data", *BIOINFORMATICS*, vol. 19, pp. 1578-1579, 2003
- [8] K. Murphy. "Machine learning: a probabilistic perspective". Cambridge MA: MIT Press, 2012.
- [9] N. Thouleimat, D. Hernandez-Lobato, and P. Dupont, "Variance Estimators for t-Test Ranking Influence the Stability and Predictive Performance of Microarray Gene Signatures", *European Conference on Computational Biology*, 2010.
- [10] S. Sahan, K. Polata, H. Kodazb, and S. Gne, "Anewhybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis", *ELSEVIER*, vol. 37, no. pp. 415-423, 2007.
- [11] M. Visser, M. Pahor, F. Tylavsky, S. B. Kritchevsky, J. A. Cauley, A. B. Newman, B. A. Blunt, and T. B. Harris, "One-and two-year change in body composition as measured by DXA in a population-based cohort of older men and women", *Journal of applied physiology*, vol. 94, no. pp. 2368-2374, 2003.
- [12] V. A. Hughes, W. R. Frontera, R. Roubenoff, W. J. Evans, and M. A. F. Singh, "Longitudinal changes in body composition in older men and women: role of body weight change and physical activity", *The American journal of clinical nutrition*, no. pp. 473-481, 2002
- [13] I. Janssen, S. B. Heymsfield, Z. Wang, and R. Ross, "Skeletal muscle mass and distribution in 468 men and women aged 1888 yr", *Journal of applied physiology* vol. 89.no.1, pp. 81-88, 2000.
- [14] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, and S. Y. Philip, "Top 10 algorithms in data mining", *Knowledge and Information Systems*, vol. 14, no. pp. 1-37, 2008.
- [15] A. C. Haury, P. Gestraud, and J.-P. Vert, "The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures", *PloS one*, vol. 6, p. e28210, 2011.
- [16] W. Altidor, T. M. Khoshgoftaar and J V Hulse and A. Napolitano, "Ensemble Feature Ranking Methods for Data Intensive Computing Applications", *Handbook of Data Intensive Computing*, pp 349-376, 2011
- [17] A. Y. Guo, K. S. LeunG, P. M. F. Siu, J. H. Qin, S. K. H. Chow, L. Qin, C. Y. Li, and W. H. Cheung, "Muscle mass, structural and functional investigations of senescence-accelerated mouse P8 (SAMP8)", *Experimental Animals*, vol. 64, p. 425, 2015.
- [18] R. R. Kalyani, M. Corriere, and L. Ferrucci, "Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases", *The Lancet Diabetes & Endocrinology*, vol. 2, no. pp. 819-829, 2014.

- [19] U. Raue, T. A. Trappe, S. T. Estrem, H.-R. Qian, L. M. Helvering, R. C. Smith, and S. Trappe, "Transcriptome signature of resistance exercise adaptations: mixed muscle and fiber type specific profiles in young and old adults," *Journal of Applied Physiology*, vol. 112, no. pp. 1625-1636, 2012.
- [20]. R. Sharma, R. B. Pachori, and U. R. Acharya, "An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures," *Entropy*, vol. 17, no. pp. 5218-5240, 2015.
- [21] U. R. Acharya, E. Ng, L. W. J. Eugene, K. P. Noronha, L. C. Min, K. P. Nayak, and S. V. Bhandary, "Decision support system for the glaucoma using Gabor transformation," *Biomedical Signal Processing and Control*, vol. 15, no. pp. 18-26, 2015.